Conditional Functional Graphical Models

Kuang-Yao Lee[a], **Dingjue Ji**[b], **Lexin Li**[c], **Todd Constable**[b,d], **Hongyu Zhao**[b]

[a]Department of Statistical Science, Temple University, Philadelphia, PA;

[b]Department of Biostatistics, Yale University, New Haven, CT;

[c]Division of Biostatistics, University of California, Berkeley, CA;

[d]Department of Radiology and Biomedical Imaging, Yale University, New Haven, CT

Abstract

Graphical modeling of multivariate functional data is becoming increasingly important in a wide variety of applications. The changes of graph structure can often be attributed to external variables, such as the diagnosis status or time, the latter of which gives rise to the problem of dynamic graphical modeling. Most existing methods focus on estimating the graph by aggregating samples, but largely ignore the subject-level heterogeneity due to the external variables. In this article, we introduce a conditional graphical model for multivariate random functions, where we treat the external variables as conditioning set, and allow the graph structure to vary with the external variables. Our method is built on two new linear operators, the conditional precision operator and the conditional partial correlation operator, which extend the precision matrix and the partial correlation matrix to both the conditional and functional settings. We show that their nonzero elements can be used to characterize the conditional graphs, and develop the corresponding estimators. We establish the uniform convergence of the proposed estimators and the consistency of the estimated graph, while allowing the graph size to grow with the sample size, and accommodating both completely and partially observed data. We demonstrate the efficacy of the method through both simulations and a study of brain functional connectivity network.

Keywords

Brain connectivity analysis; Functional magnetic resonance imaging; Graphical model; Karhunen–Loève expansion; Linear operator; Reproducing kernel Hilbert space

1. Introduction

Functional graphical modeling is gaining increasing attention in the recent years, where the central goal is to investigate the interdependence among multivariate random functions. Applications include time course gene expression data in genomics (Wei and Li 2008),

multivariate time series data in finance (Tsay and Pourahmadi 2017), electrocorticography, and functional magnetic resonance data in neuroimaging (Zhang et al. 2015), among many others.

Our motivation is brain functional connectivity analysis based on functional magnetic resonance imaging (fMRI). Functional MRI measures brain neural activities via blood oxygenlevel-dependent signals. It depicts brain functional connectivity network, which is shown to alter under different disorders or during different brain developmental stages. Such alterations contain crucial insights of both disorder pathology and development of the brain (Fox and Greicius 2010). The fMRI data are often summarized in the form of a location by time matrix for each individual subject. The rows correspond to a set of brain regions, and the columns correspond to time points that are usually 500 msec to 2 sec apart and span a few minutes in total. From the fMRI scans, a graph is constructed, where nodes represent brain regions, and links represent interactions and dependencies among the regions (Fornito, Zalesky, and Breakspear 2013). Numerous statistical methods have been developed to estimate functional connectivity network. Most of these methods treat the fMRI data as multivariate random variables with repeated observations, where each region is represented by a random variable and the time-course data are taken as repeated measures of that variable (e.g., Bullmore and Sporns 2009; Wang et al. 2016b). There are recent proposals to model the fMRI data as multivariate functions, where the time-course data of each region are taken as a function (e.g., Li and Solea 2018). Given the continuous nature and the short-time interval between the adjacent sampling points of fMRI, we treat the data as multivariate functions, and formulate the connectivity network estimation as a functional graphical modeling problem in this article.

Nearly, all existing graph estimation methods tackle the problem by aggregating samples, sometimes according to the diagnostic groups. However, there is considerable subject-level heterogeneity, which may contain crucial information for our understanding of the network, but has been largely untapped or ignored by existing methods (Fornito, Zalesky, and Breakspear 2013). Heterogeneity could arise due to a subject's phenotype profile; for example, in our study in Section 7, the brain connectivity network may vary with an individual's intelligence score. It could also arise due to a time variable, for example, the subject's age, and the connectivity network may vary with age, which leads to dynamic graphical modeling. In this article, we introduce a *conditional functional graphical model* for a set of random functions, by modeling the external variables such as the phenotype or age as the conditioning set. Our proposal thus extends two lines of existing and relevant research: from conditional and dynamic graphical modeling of random variables to that of random functions, and from unconditional functional graphical modeling to conditional functional graphical modeling.

The first line of relevant research is the class of graphical models of random variables. Within this class, there is a rich literature on unconditional graphical models (Yuan and Lin 2007; Friedman et al. 2008; Peng et al. 2009, among others). There are extensions to joint estimation of multiple graphs, which arise from a small number of groups, typically two or three, according to an external variable such as the diagnostic status (Danaher, Wang, and Witten 2011; Chun, Zhang, and Zhao 2015; Lee and Liu 2015; Zhu and Li 2018).

There have also been some recent proposals of dynamic graphical models (Kolar et al. 2010; Xu and Hero 2014; Zhang and Cao 2017). However, they only considered a discrete-time setting, in which the network is estimated only at a small and discrete number of time points. The most relevant work to ours is Qiu et al. (2016), who also targeted estimation of the individual graph according to an external variable, for example, age. Although both Qiu et al. (2016) and our method are designed for graph estimation at the individual level, the two solutions differ in many ways. Particularly, Qiu et al. (2016) assumed that the repeated observations of each individual come from the same Gaussian distribution, whose dependency was required to follow a certain stationary time series structure. By contrast, we treat the repeated measurements as realizations from a random function, and do not impose any structural or relational assumption on the entire function. More importantly, compared to the random variable setting in Qiu et al. (2016), our functional setting involves an utterly different and new set of modeling techniques and theoretical tools.

The second line of relevant research is the class of unconditional graphical models of random functions, which appeared only recently. Qiao, Guo, and James (2019) introduced a functional Gaussian graphical model, by assuming that the random functions follow a multivariate Gaussian distribution. Li and Solea (2018) relaxed the Gaussian assumption, developed a precision operator that generalizes the concept of precision matrix to the functional setting, and used it to estimate the unconditional functional graph. However, their precision operator is nonrandom, and their graph dimension is fixed. By contrast, we introduce a conditional precision operator (CPO), which is a function of the conditioning variable and is thus random, and we allow the graph dimension to diverge with the sample size. These differences bring extra challenges in analyzing the operator-based statistics. Moreover, because the relation between the CPO and the conditioning variable can be nonlinear, its estimation requires the construction of a reproducing kernel Hilbert space (RKHS) of the conditioning variable, which leads to a more complex asymptotic analysis than that of Li and Solea (2018). We also derive a number of concentration bounds and uniform convergence for our proposed estimators, while such results are not available in Li and Solea (2018).

To address the problem of conditional graphical modeling of random functions, we introduce two new linear operators: the CPO, and the conditional partial correlation operator (CPCO), which extend the precision matrix and the partial correlation matrix from the random variable setting to both the conditional and functional settings. We show that, when the conditional distribution is Gaussian, the conditional graph can be thoroughly captured by the nonzero elements of CPO or CPCO. This property echoes the classic result where a static graph can be inferred by the precision matrix or the partial correlation matrix under the Gaussian assumption. We note that, some early works, such as Lee, Li, and Zhao (2016a,b), also estimated the parameters of interest through linear operators. However, we studied utterly different problems: Lee, Li, and Zhao (2016b) targeted variable selection in classical regressions, Lee, Li, and Zhao (2016a) targeted unconditional graph estimation for random variables, while we target conditional graph estimation for random functions. Both the methodology and theory involved are thus substantially different.

Our proposal makes useful contributions at multiple fronts. On the method side, it offers a new class of statistical models to study conditional graph estimation for multivariate functional data, a problem that remains largely unaddressed. We investigate the parallels between the random variable-based and random function-based graphs, and between the unconditional and conditional graphs. On the theory side, our work develops new tools for operator-based functional data analysis. We establish the conditional graph estimation consistency, along with a set of concentration inequalities and error bounds, for our proposed method. To our knowledge, very little work has investigated function-on-function dependency at such a level of complexity that involves estimating the linear operators under a conditional framework. The tools we develop are general, and can be applied to other settings in high-dimensional functional data analysis. On the computation side, under a properly defined coordinate system, the proposed operators are functions of the sample covariance operator of dimension $n \times n$, with n being the sample size. It is relatively easy to calculate, and the accompanying estimation algorithm can be scaled up to large graphs.

The rest of the article is organized as follows. We begin with a formal definition of conditional functional graphical model in Section 2. We introduce a series of linear operators in Section 3, develop their estimators in Section 4, and study their asymptotic properties in Section 5. We report the simulations in Section 6, and an analysis of an fMRI dataset in Section 7. We relegate all proofs and some additional results to the supplementary appendix.

2. Model

In this section, we formally define the conditional functional graphical model. Let $(\Omega_X, \mathscr{F}_X)$ be a measurable space. Suppose Ω_{X_i} is a Hilbert space of $\mathbb{R}$-valued functions on an interval $T \subset \mathbb{R}$, for $i = 1, \ldots, p$, and Ω_X is the Cartesian product $\Omega_{X_1} \times \cdots \times \Omega_{X_p}$. Suppose $X = (X_1, \ldots, X_p)^\top$ is a p-dimensional random element on Ω_X. Let $G = (V, E)$ be an undirected graph, where $V = \{1, \ldots, p\}$ represents the set of vertices corresponding to the p random functions, and $E = \{(i, j) \in V \times V, i = j\}$ represents the set of undirected edges. A common approach to modeling an undirected graph is to associate the separation on the graph with the conditional independence; in other words, nodes i and j are separated in G if and only if X_i and X_j are independent given the rest of X, that is,

$$(i, j) \notin E \quad \Leftrightarrow \quad X_i \perp\!\!\!\perp X_j | X_{-(i, j)}, \tag{1}$$

where $X_{-(i,j)}$ represents X with its ith and jth components removed, and $\perp\!\!\!\perp$ represents statistical independence. Based on Equation (1), Qiao, Guo, and James (2019) proposed a functional graphical model, which assumed that X follows a multivariate Gaussian distribution.

Next, we introduce a conditional functional graphical model that allows the graph links to vary with the external variables. We focus on the univariate external variable case, but our method can be generalized to multivariate external variables, or external functions. Let Y be a random element defined on Ω_Y, and $\mathscr{F}_Y$ be the Borel σ-field generated by the open sets in Ω_Y. Let $\mathscr{F}_Y \to \mathbb{R}$ and $P_{X|Y}(\cdot \mid \cdot): \mathscr{F}_X \times \Omega_Y \to \mathbb{R}$ be the distribution of Y and conditional distribution of $X | Y$, respectively. We next give our formal definition.

Definition 1.

Suppose a random graph E^y for each $y \in \Omega_Y$ is defined via the mapping $\Omega_Y \to 2^{V \times V}$, $y \to E^y$, where $2^{V \times V}$ is the power set of $V \times V$. We say X follows a conditional functional graphical model with respect to E^y if and only if, for $y \in \Omega_Y$,

$$(i, j) \notin E^y \quad \Leftrightarrow \quad X_i \perp\!\!\!\perp X_j \mid [X_{-(i,j)}, Y = y]. \tag{2}$$

We note that Li, Chun, and Zhao (2012) introduced the notion of conditional graphical model. However, our notion of conditional functional graphical model is considerably different, in that our model extends theirs not only from the setting of random variables to random functions, but also from the setting of static graphs to random graphs. Specifically, letting $X = (X_1, \ldots, X_p)^\top$ and $Y = (Y_1, \ldots, Y_q)^\top$ denote two random vectors, Li, Chun, and Zhao (2012) considered the model,

$$(i, j) \notin E^0 \quad \Leftrightarrow \quad X_i \perp\!\!\!\perp X_j \mid [X_{-(i,j)}, Y = y]$$

for all $y \in \mathbb{R}^q$. In this model, $E^0 \subseteq 2^{V \times V}$ is a fixed graph, and does not change with the value of Y. In comparison, our model in Equation (2) allows X to be a p-variate random function, and the graph E^y to vary with Y.

3. Linear Operators

In this section, we first introduce a series of linear operators, based on which we then formally define the CPO and the CPCO. Finally, we study the relation between these two operators and the conditional functional graph.

We adopt the following notation throughout this article. For two generic Hilbert spaces Ω, Ω', let $\mathcal{B}(\Omega, \Omega')$ and $\mathcal{B}_2(\Omega, \Omega')$ denote the class of all bounded and Hilbert–Schmidt operators from Ω to Ω', respectively. We abbreviate $\mathcal{B}(\Omega, \Omega)$ and $\mathcal{B}_2(\Omega, \Omega)$ as $\mathcal{B}(\Omega)$ and $\mathcal{B}_2(\Omega)$ whenever they are appropriate. Moreover, let $\|\cdot\|$ and $\|\cdot\|_{HS}$ be the operator norm in $\mathcal{B}(\Omega)$ and Hilbert-Schmidt norm in $\mathcal{B}_2(\Omega)$. Moreover, let $\ker(A)$, $\mathrm{range}(A)$, $\overline{\mathrm{range}}(A)$ denote the null space, the range, and the closure of the range of an operator A, respectively.

3.1. Conditional Covariance and Correlation Operators

We first define three covariance operators, $\mathfrak{B}_{YY}$, $\mathfrak{B}_{X_iX_j}$, and $\mathfrak{B}_{YX_{ij}}$. We then define the conditional covariance operator $\mathfrak{B}^y_{X_iX_j}$, and the conditional correlation operator $\mathfrak{C}^y_{X_iX_j}$, which is the building block of CPO and CPCO.

Let $\kappa_Y : \Omega_Y \times \Omega_Y \to \mathbb{R}$ be a positive-definite kernel, $\mathscr{H}_Y$ be its corresponding RKHS, and $L_2(P_Y)$ be the collection of all square-integrable functions of Y under P_Y. The next assumption ensures the square-integrability of X_i under $P_{X|Y}$, and that $\mathscr{H}_Y$ is a subset of $L_2(P_Y)$.

Assumption 1.—There exist $M_0 > 0$ and $M_Y > 0$ such that
$\sup\{E(\|X_i\|^2_{\Omega_{X_i}} \mid y): y \in \Omega_Y\} \le M_0$, for $i = 1, \ldots, p$, and $\kappa_y(Y, Y) \quad M_Y$.

We comment that, we choose RKHS as the modeling space, so that the relation of X on Y can be very flexible, and the kernel matrix of Y is of dimension $n \times n$, an attractive feature when the dimension of Y is large compared with n. In addition, a good number of asymptotic tools for RKHS operators have been developed (see, Bach 2009; Lee, Li, and Zhao 2016a). That being said, our theoretical development can also be easily extended to the spaces beyond RKHS. In fact, our population development only requires that $\mathcal{H}_Y$ is a proper subset of $L_2(P_Y)$, which can be ensured by the square-integrable condition, $\mathrm{var}[h(Y)] = M \| h \|_{\mathcal{H}_Y}$ for an $M > 0$ and every $h \in \mathcal{H}_Y$.

Let $\otimes$ denote the tensor product; then $\kappa_y(\cdot, Y) \otimes \kappa_y(\cdot, Y)$, $X_i \otimes X_j$ and $\kappa_y(\cdot, Y) \otimes X_i \otimes X_j$ are random elements in $\mathcal{B}_2(\mathcal{H}_Y)$, $\mathcal{B}_2(\Omega_{X_j}, \Omega_{X_i})$, and $\mathcal{B}_2[\mathcal{B}_2(\Omega_{X_j}, \Omega_{X_i}), \mathcal{H}_Y]$, respectively. Their expectations uniquely define the covariance operators,

$$\begin{aligned}
&\mathfrak{B}_{YY} = E[\kappa_y(\cdot, Y) \otimes \kappa_y(\cdot, Y)], \text{ via} \\
&\langle h_1, \mathfrak{B}_{YY} h_2 \rangle = E[h_1(Y) h_2(Y)], \\
&\mathfrak{B}_{X_i X_j} = E(X_i \otimes X_j), \text{ via} \\
&\langle f, \mathfrak{B}_{X_i X_j} g \rangle = E(\langle f, X_i \rangle \langle g, X_j \rangle), \\
&\mathfrak{B}_{YX_{ij}} = E[\kappa_y(\cdot, Y) \otimes X_i \otimes X_j], \text{ via} \\
&\langle h, \mathfrak{B}_{YX_{ij}}(f \otimes g) \rangle = E[\langle X_i, f \rangle \langle X_j, g \rangle h(Y)],
\end{aligned} \tag{3}$$

for all $f \in \Omega_{X_j}$, $g \in \Omega_{X_i}$, and $h, h_1, h_2 \in \mathcal{H}_Y$. The next proposition justifies the existence of $\mathfrak{B}_{YY}$, $\mathfrak{B}_{X_i X_j}$, and $\mathfrak{B}_{YX_{ij}}$.

Proposition 1.—If Assumption 1 holds, then there exist linear operators $\mathfrak{B}_{YY}$, $\mathfrak{B}_{X_i X_j}$, and $\mathfrak{B}_{YX_{ij}}$ satisfying the relations in Equation (3).

We next introduce a regression operator, $\mathfrak{M}_{X_{ij} \mid Y}$ through the relation, $\mathfrak{M}_{X_{ij} \mid Y} = \mathfrak{B}_{YY}^\dagger \mathfrak{B}_{YX_{ij}}$, where $\dagger$ is the Moore–Penrose inverse. We first need an assumption on the ranges of $\mathfrak{B}_{YY}$ and $\mathfrak{B}_{YX_{ij}}$, and $\mathcal{H}_Y$ is sufficiently rich in $L_2(P_Y)$.

Assumption 2.—For every $(i, j) \in V \times V$, $\mathrm{range}(\mathfrak{B}_{YX_{ij}}) \subseteq \mathrm{range}(\mathfrak{B}_{YY})$. Moreover, $\mathcal{H}_Y$ is dense in $L_2(P_Y)$.

By Assumption 2, for any $h \in \mathrm{range}(\mathfrak{B}_{YY})$, there exists a unique $h' \in \overline{\mathrm{range}}(\mathfrak{B}_{YY})$ such that $h = \mathfrak{B}_{YY} h'$. Therefore, the inverse $\mathfrak{B}_{YY}^\dagger$ is defined as $\mathfrak{B}_{YY}^\dagger: \mathrm{range}(\mathfrak{B}_{YY}) \to \overline{\mathrm{range}}(\mathfrak{B}_{YY})$, $h \mapsto h'$, which implies that $\mathfrak{M}_{X_{ij} \mid Y}$ is well-defined. The range condition that $\mathrm{range}(\mathfrak{B}_{YX_{ij}}) \subseteq \mathrm{range}(\mathfrak{B}_{YY})$ is satisfied generally. For instance, it holds when the rank of $\mathcal{B}_2(\Omega_{X_j}, \Omega_{X_i})$ is finite, which is reasonable, because in practice Ω_{X_i} can often be approximated by the spanning space of a few leading eigenfunctions of $\mathfrak{B}_{X_i X_i}$.

The next proposition shows that $\mathfrak{M}_{X_{ij}|Y}$ maps every $(f, g) \in \Omega_{X_i} \times \Omega_{X_j}$ to $E(\langle f, X_i \rangle \langle g, X_j \rangle \mid y)$, that is, the conditional expectation, and thus this operator can be viewed as a regression operator.

Proposition 2.—If Assumptions 1 and 2 hold, then for all $y \in \Omega_Y$ and $(f, g) \in \Omega_{X_i} \times \Omega_{X_j}$, we have $\left[\mathfrak{M}_{X_{ij}|Y}(f \otimes g) \right] \circ (y) = E(\langle f, X_i \rangle \langle g, X_j \rangle \mid y)$.

Now we are ready to define the conditional covariance operator, whose existence is justified by Proposition 2 and Riesz representation theorem.

Definition 2.—For each $y \in \Omega_Y$, the bilinear form, $\Omega_{X_i} \times \Omega_{X_j} \to \mathbb{R}$, $(f, g) \mapsto \mathfrak{M}_{X_{ij}|Y}(f \otimes g) \circ (y)$, uniquely defines an operator $\mathfrak{V}_{X_i X_j}^y \in \mathscr{B}(\Omega_{X_j}, \Omega_{X_i})$, via $\langle f, \mathfrak{V}_{X_i X_j}^y g \rangle_{\Omega_{X_i}} = E(\langle f, X_i \rangle \langle g, X_j \rangle \mid y)$ for all $(f, g) \in \Omega_{X_i} \times \Omega_{X_j}$. We call $\mathfrak{V}_{X_i X_j}^y$ the conditional covariance operator.

Note that the mapping $\Omega_Y \to \mathscr{B}(\Omega_{X_j}, \Omega_{X_i})$, $y \mapsto \mathfrak{V}_{X_i X_j}^y$ defines a random operator. If the conditional expectations $E(\langle f, X_i \rangle \mid y) = 0$, then $\mathfrak{V}_{X_i X_i}^y$ induces the conditional covariance $\text{cov}(\langle f, X_i \rangle, \langle g, X_j \rangle \mid y)$. When X_i and X_j are random vectors, Fukumizu, Bach, and Jordan (2009) introduced the homoscedastic conditional covariance operator, which induces $E[\text{cov}(\langle f, X_i \rangle, \langle g, X_j \rangle \mid Y)]$. Our conditional covariance operator is different from that of Fukumizu, Bach, and Jordan (2009), as it extends the classical conditional covariance to the functional setting, and it deals directly with $\text{cov}(\langle f, X_i \rangle, \langle g, X_j \rangle \mid y)$, instead of its expectation. We write the joint operator $\mathfrak{V}_{XX}^y : \Omega_X \to \Omega_X$ as the block matrix of operators whose (i, j)th element is $\mathfrak{V}_{X_i X_j}^y, 1 \leq i, j \leq p$, or more explicitly, $\mathfrak{V}_{XX}^y f = \left(\sum_{j=1}^p \mathfrak{V}_{X_1 X_j} f_j, ..., \sum_{j=1}^p \mathfrak{V}_{X_p X_j} f_j \right)^\top$, for any $f = (f_1, ..., f_p)^\top \in \Omega_X$.

Given the conditional covariance operator $\mathfrak{V}_{X_i X_j}^y$, we next define the conditional correlation operator, $\mathfrak{C}_{X_i X_j}^y : \Omega_{X_j} \to \Omega_{X_i}$, for each $y \in \Omega_Y$ via,

$$\mathfrak{V}_{X_i X_j}^y = \left(\mathfrak{V}_{X_i X_i}^y \right)^{1/2} \mathfrak{C}_{X_i X_j}^y \left(\mathfrak{V}_{X_j X_j}^y \right)^{1/2}, \text{ with } \| \mathfrak{C}_{X_i X_j}^y \| \leq 1. \tag{4}$$

Its existence and uniqueness are ensured by Baker (1973). Similar to the construction of $\mathfrak{V}_{XX}^y$, we write the joint operator $\mathfrak{C}_{XX}^y : \Omega_X \to \Omega_X$ as the block matrix of operators whose (i, j)th element is $\mathfrak{C}_{X_i X_j}^y, 1 \leq i, j \leq p$. Let $\mathfrak{D}_X^y$ denote the block diagonal matrix $[\mathfrak{D}_X^y]_{i,i} = \mathfrak{V}_{X_i X_i}^y$, for $i \in V$. We then have $\mathfrak{V}_{XX}^y = [\mathfrak{D}_X^y]^{1/2} \mathfrak{C}_{XX}^y [\mathfrak{D}_X^y]^{1/2}$.

Next, we impose the distributional assumption on $X \mid y$.

Assumption 3.—Suppose $X \mid y$ follows the conditional functional graphical model as defined via (2), and the conditional distribution of $X = (X_1, ..., X_p)^\top$ given $Y = y$ follows a centered Gaussian distribution.

The zero-mean condition is imposed to simplify both methodological and theoretical development, and can be relaxed with some modifications. Moreover, we require the

conditional distribution $X \mid y$ to follow a Gaussian distribution. It is possible to relax this Gaussian assumption, by extending the notion of functional additive conditional independence (Li and Solea 2018), or the copula graphical model (Liu et al. 2012). However, we feel that the Gaussian case itself is worthy of a full investigation, and we leave the non-Gaussian extension as future research.

Assumption 3, together with Definition 2, imply that, for $f = (f_1, \ldots, f_p)^\top \in \Omega_X$,
$E\big(\exp \sum_{i=1}^p \iota \langle f_i, X_i \rangle_{\Omega_{X_i}} \mid y\big) = \exp\big(-1/2 \sum_{i,j=1}^p \langle f_i, \mathfrak{B}^y_{X_i X_j} f_j \rangle_{\Omega_{X_i}}\big)$, where $\iota = \sqrt{-1}$.

3.2. Conditional Precision and Partial CorrelationOperators

We next formally define the operators CPO and CPCO, then establish their relations with the conditional functional graph. We introduce two assumptions, which ensure that $\mathfrak{C}^y_{X_i X_j}$ is Hilbert-Schmidt and $\mathfrak{C}^y_{XX}$ is invertible. We present some intuitions here, but relegate the detailed technical discussion to Section S.3 in the appendix.

Assumption 4.—For each $y \in \Omega_Y$ and $i \in V$, let $\{(\lambda^{y,a}_i, \eta^{y,a}_i)\}_{a \in \mathbb{N}}$ denote the collection of eigenvalue and eigenfunction pairs of $\mathfrak{B}^y_{X_i X_i}$. Let $\mathbb{N}^y_i = \{a \in \mathbb{N} : \lambda^{y,a}_i > 0\}$. There exists $c_1 > 0$ such that,

$$
\begin{aligned}
&\max_{i,j \in V, i \neq j} \sum_{a \in \mathbb{N}^y_i, b \in \mathbb{N}^y_j} \frac{\mathrm{cov}^2\big(\langle \eta^{y,a}_i, X_i \rangle, \langle \eta^{y,b}_j, X_j \rangle \mid y\big)}{\mathrm{var}^2\big(\langle \eta^{y,a}_i, X_i \rangle \mid y\big)\mathrm{var}^2\big(\langle \eta^{y,b}_j, X_j \rangle \mid y\big)} \\
&\equiv \max_{i,j \in V, i \neq j} \sum_{a \in \mathbb{N}^y_i, b \in \mathbb{N}^y_j} \big(\rho^{y,a,b}_{i,j}\big)^2 \leq c_1 .
\end{aligned}
\tag{5}
$$

Assumption 5.—For each $y \in \Omega_Y$, $\ker(\mathfrak{B}^y_{XX}) = 0$.

Assumption 4 characterizes the level of smoothness for the underlying distributions of the random functions. Assumption 5 is to prevent the existence of a constant function consisting of linear combination of nonconstant functions. It can be viewed as the generalization of the nonexistence of collinearity in linear models, or the empty concavity space in generalized liner models (Hastie and Tibshirani 1990). Assumption 4 ensures that $\mathfrak{C}^y_{X_i X_j}$ is Hilbert-Schmidt, and thus compact. Mean-while, Assumptions 4 and 5 together ensure that $\mathfrak{C}^y_{XX}$ is lower-bounded by a strictly positive constant. This implies that $\mathfrak{C}^y_{XX}$ is invertible, and that $\mathfrak{P}^y = [\mathfrak{C}^y_{XX}]^{-1}$ is bounded, which justifies the following definition.

Definition 3.—Define the CPO as the inverse of the joint conditional correlation operator, $\mathfrak{P}^y = [\mathfrak{C}^y_{XX}]^{-1} \in \mathscr{B}(\Omega_X)$, for any $y \in \Omega_Y$.

The operator $\mathfrak{P}^y$ generalizes the precision matrix to the functional and conditional settings. We should clarify that, unlike the standard definition where the precision matrix is defined as the inverse of the covariance matrix (Cai, Liu, and Luo 2011), our CPO is defined as the inverse of the conditional correlation operator. This is to avoid taking inversion on the covariance operator, which is usually not invertible because of its compactness. Next, we develop an operator that generalizes the partial correlation matrix to the functional

and conditional setting. Similar to the definition for $\mathfrak{B}^y_{XX}$, we define $\mathfrak{B}^y_{X_A X_B}$ for any subsets A, B $\subseteq$ V. Therefore, for any subsets A $\subseteq$ V\\$\{i, j\}$, We define an intermediate operator, $\mathfrak{B}^y_{X_i X_j | X_A} : \Omega_{X_j} \to \Omega_{X_i}$ through the relation, $\mathfrak{B}^y_{X_i X_j | X_A} = \mathfrak{B}^y_{X_i X_j} - \mathfrak{B}^y_{X_i X_A}[\mathfrak{B}^y_{X_A X_A}]^\dagger \mathfrak{B}^y_{X_A X_j}$, for any $(i, j) \in$ V $\times$ V. We then have the following result if A$= -(i, j)$.

Proposition 3.—There uniquely exists $\mathfrak{R}^y_{X_i X_j | X_{-(ij)}} \in \mathscr{B}(\Omega_{X_j}, \Omega_{X_i})$ which satisfies that $\mathfrak{B}^y_{X_i X_j | X_{-(i,j)}} = [\mathfrak{B}^y_{X_i X_i | X_{-(i,j)}}]^{1/2} \mathfrak{R}^y_{X_i X_j | X_{-(i,j)}} [\mathfrak{B}^y_{X_j X_j | X_{-(i,j)}}]^{1/2}$, and $\|\mathfrak{R}^y_{X_i X_j | X_{-(i,j)}}\| \le 1$.

Its proof is similar to that of (Lee, Li, and Zhao 2016a, theo. 1) and is thus omitted. It justifies the definition of the following operator.

Definition 4.—We call the operator $\mathfrak{R}^y_{X_i X_j | X_{-(i,j)}}$ in Proposition 3 the CPCO between X_i and X_j given $X_{-(i,j)}$ and Y.

3.3. Relation With Conditional Functional Graph

We first show that the conditional covariance operator can be constructed by the functions of conditional covariances between the Karhunen-Loève coefficients and the associated eigenfunctions. This simple form provides a convenient way to estimate the conditional covariance operator later. Let $\{(\lambda_i^a, \eta_i^a)\}_{a \in \mathbb{N}}$ denote the collection of eigenvalue and eigenvector pairs of $\mathfrak{B}_{X_i X_i}$, with $\lambda_i^1 \ge \lambda_i^2 \ge \cdots \ge 0$. Then X_i can almost surely be represented as $X_i = \sum_{a \in \mathbb{N}} \alpha_i^a \eta_i^a$, where $\alpha_i^a = \langle X_i, \eta_i^a \rangle$, for all $a \in \mathbb{N}$. This expression is known as the Karhunen-Loève (K-L) expansion (Bosq 2000).

Proposition 4.—Suppose the same conditions in Proposition 2 hold. Then we have

$$\mathfrak{B}^y_{X_i X_j} = \sum_{a, b \in \mathbb{N}} E(\alpha_i^a \alpha_j^b | y)(\eta_i^a \otimes \eta_j^b), \text{ for each } y \in \Omega_Y,$$

where (α_i^a, η_i^a) and (α_j^b, η_j^b) are from the Karhunen-Loève expansion.

We next show that CPO and CPCO fully characterize the conditional functional independence, and are thus crucial for our estimation of conditional functional graph.

Theorem 1.—If Assumptions 1–5 hold, then we have, for any $y \in \Omega_Y$,

$$X_i \perp\!\!\!\perp X_j | [X_{-(i,j)}, Y = y] \quad \Leftrightarrow \quad [\mathfrak{P}^y]_{i,j} = 0,$$

where $[\mathfrak{P}^y]_{i,j}$ denotes the (i, j)th element of $\mathfrak{P}^y$.

Under the Gaussian distribution, the equivalence between the conditional independence and the zero element of a nonrandom precision matrix is well known in the classical random variables setting. By contrast, Theorem 1 extends to the setting of random functions and also allows the precision operator to vary with Y.

Theorem 2.—If Assumptions 1–3 hold, then we have, for any $y \in \Omega_Y$,

$$X_i \perp\!\!\!\perp X_j \mid [X_{-(i,j)}, Y = y] \quad \Leftrightarrow \quad \mathfrak{R}^y_{X_i X_j \mid X_{-(i,j)}} = 0.$$

Theorems 1 and 2 suggest that one can estimate the conditional functional graph E^y in Equation (2) through the proposed operators, CPO or CPCO. In the following, we primarily focus on the graph estimation based on CPO, and investigate the corresponding asymptotics. The results based on CPCO can be derived in a parallel fashion, which we only briefly discuss in Section S.4 in the Appendix.

4. Estimation

In this section, we first derive the sample estimate of CPO and the conditional graph at the operator level. We then construct empirical bases and develop coordinate representations for the functions observed at a finite set of time points. Using these coordinate representations, we are able to compute our estimated linear operators. Last, we provide a step-by-step summary of our proposed estimation procedure.

4.1. Operator-Level Estimation

We first derive the sample Karhunen-Loève expansion. We and then sequentially develop the estimators of $\mathfrak{B}^y_{XX}$, $\mathfrak{C}^y_{XX}$, $\mathfrak{P}^y$, and finally E^y.

Let $(Y^1, \ldots, Y^n)$ denote iid samples of Y, and $(X^1, \ldots, X^n)$ denote iid samples of X, with $X^k = (X^k_1, \ldots, X^k_p)^\top$, for $k = 1, \ldots, n$. Let E_n denote the sample mean operator; that is, for a sample $(\omega^1, \ldots, \omega^n)$ from Ω, $E_n(\omega) = \sum^n_{k=1} \omega^k / n$. We estimate the covariance operators, $\mathfrak{B}_{X_i X_i}$, $\mathfrak{B}_{Y X_{ij}}$, and $\mathfrak{B}_{YY}$, by

$$\widehat{\mathfrak{B}}_{X_i X_i} = E_n(X_i \otimes X_i),$$

$$\widehat{\mathfrak{B}}_{Y X_{ij}} = E_n[\kappa_Y(\cdot, Y) \otimes X_i \otimes X_j],$$

$$\widehat{\mathfrak{B}}_{YY} = E_n[\kappa_Y(\cdot, Y) \otimes \kappa_Y(\cdot, Y)],$$

for any $(i, j) \in \mathrm{V} \times \mathrm{V}$. For $i \in \mathrm{V}$, let $\left\{ \left(\hat{\lambda}^a_i, \hat{\eta}^a_i \right) \right\}_{a \in \mathbb{N}}$ denote the collection of eigenvalue and eigenfunction pairs of $\widehat{\mathfrak{B}}_{X_i X_i}$. Then, we have $X^k_i = \sum_{a \in \mathbb{N}} \hat{\alpha}^{k,a}_i \hat{\eta}^a_i$, where $\hat{\alpha}^{k,a}_i = \langle X^k_i, \hat{\eta}^a_i \rangle$ is the ath coefficient from the K-L expansion of X^k_i for the kth subject. Furthermore, we use the leading d terms to approximate X^k_i; in other words, we have, $X^k_i \approx \sum^d_{a=1} \hat{\alpha}^{k,a}_i \hat{\eta}^a_i$, for all $k = 1, \ldots, n$, and $i \in \mathrm{V}$.

By its definition, we estimate the regression operator $\mathfrak{M}_{X_{ij} \mid Y}$ by

$$\widehat{\mathfrak{M}}_{X_{ij}|Y}(\epsilon_Y) = \left(\widehat{\mathfrak{B}}_{YY} + \epsilon_Y I\right)^{-1}\widehat{\mathfrak{B}}_{YX_{ij}}, \tag{6}$$

where $\epsilon_Y > 0$ is a prespecified ridge parameter, and it imposes a level of smoothness on the regression structure. Next, by Propositions 2 and 4, given y, d, i, j, we estimate the conditional covariance operator $\mathfrak{B}_{X_iX_j}^y$ by

$$\widehat{\mathfrak{B}}_{X_iX_j}^y(d, \epsilon_Y) = \sum_{a,b=1}^{d} \left\{ \left[\widehat{\mathfrak{M}}_{X_{ij}|Y}(\epsilon_Y)\right]\left(\widehat{\eta}_i^a \otimes \widehat{\eta}_j^b\right) \circ (y)\left(\widehat{\eta}_i^a \otimes \widehat{\eta}_j^b\right)\right\}, \tag{7}$$

where $\widehat{\eta}_i^a$ and $\widehat{\eta}_j^b$ are from the sample Karhunen-Loève expansion. Let $\widehat{\mathfrak{B}}_{XX}^y(d, \epsilon_Y)$ denote the block matrix of operators $\widehat{\mathfrak{B}}_{XX}^y(d, \epsilon_Y) = \left\{\widehat{\mathfrak{B}}_{X_iX_j}^y(d, \epsilon_Y)\right\}_{i,j \in V}$. Similarly, we can define $\widehat{\mathfrak{B}}_{X_AX_B}^y(d, \epsilon_Y)$, for any A, B $\subseteq$ V.

Therefore, by Equation (4), we estimate the conditional correlation operator $\mathfrak{C}_{X_iX_j}^y$ by

$$\widehat{\mathfrak{C}}_{X_iX_j}^y(d, \epsilon_Y, \epsilon_1) = \left[\widehat{\mathfrak{B}}_{X_iX_i}^y(d, \epsilon_Y) + \epsilon_1 I\right]^{-1/2}\widehat{\mathfrak{B}}_{X_iX_j}^y(d, \epsilon_Y)\left[\widehat{\mathfrak{B}}_{X_jX_j}^y(d, \epsilon_Y) + \epsilon_1 I\right]^{-1/2}, \tag{8}$$

for $i \ne j$, $i, j \in V$, and $\widehat{\mathfrak{C}}_{X_iX_i}^y = I$, for $i \in V$, which is the identity mapping from Ω_{X_i} to Ω_{X_i}, and ϵ_1 is a ridge regularization parameter imposed on the inverses of $\widehat{\mathfrak{B}}_{X_iX_i}^y$ and $\widehat{\mathfrak{B}}_{X_jX_j}^y$. Let $\widehat{\mathfrak{C}}_{XX}^y(d, \epsilon_Y, \epsilon_1)$ denote the block matrix of operators $\widehat{\mathfrak{C}}_{XX}^y(d, \epsilon_Y, \epsilon_1) = \left[\widehat{\mathfrak{C}}_{X_iX_j}^y(d, \epsilon_Y, \epsilon_1)\right]_{i,j \in V}$. The next result shows that the norm of $\widehat{\mathfrak{C}}_{X_iX_j}^y(d, \epsilon_Y, \epsilon_1)$ is bounded by one, which resembles the property of the correlation in the classical setting.

Proposition 5.—If $\kappa_Y(y_1, y_2) \ge 0$ for any $y_1, y_2 \in \Omega_Y$, then $\left\|\left[\widehat{\mathfrak{C}}_{XX}^y(d, \epsilon_Y, \epsilon_1)\right]_{i,j}\right\| \le 1$, for any $(i, j) \in V \times V$.

Finally, we estimate the conditional precision matrix operator $\mathfrak{P}^y$ by

$$\widehat{\mathfrak{P}}^y(d, \epsilon_Y, \epsilon_1, \epsilon_2) = \left[\widehat{\mathfrak{C}}_{XX}^y(d, \epsilon_Y, \epsilon_1) + \epsilon_2 I\right]^{-1}, \tag{9}$$

where $\epsilon_2 > 0$ is another ridge parameter. Using $\widehat{\mathfrak{P}}^y(d, \epsilon_Y, \epsilon_1, \epsilon_2)$, for each $y \in \Omega_Y$, we estimate the graph E^y by

$$\begin{aligned}&\widehat{E}_{CPO}^y(d, \epsilon_Y, \epsilon_1, \epsilon_2, \rho_{CPO})\\&= \left\{(i, j) \in V \times V : \left\|\left[\widehat{\mathfrak{P}}^y(d, \epsilon_Y, \epsilon_1, \epsilon_2)\right]_{i,j}\right\|_{HS} > \rho_{CPO}, i \ne j\right\},\end{aligned} \tag{10}$$

where $\rho_{CPO} > 0$ is a thresholding parameter. That is, we can obtain an estimator of the conditional graph by hard thresholding. For notational simplicity, hereafter we abbreviate $\widehat{\mathfrak{M}}_{X_{ij}|Y}(\epsilon_Y)$, $\widehat{\mathfrak{B}}_{X_iX_j}^y(d, \epsilon_Y)$, $\widehat{\mathfrak{B}}_{XX}^y(d, \epsilon_Y)$, $\widehat{\mathfrak{C}}_{X_iX_j}^y(d, \epsilon_Y, \epsilon_1)$, $\widehat{\mathfrak{C}}_{XX}^y(d, \epsilon_Y, \epsilon_1)$, $\widehat{\mathfrak{P}}^y(d, \epsilon_Y, \epsilon_1, \epsilon_2)$, and $\widehat{E}_{CPO}^y(d, \epsilon_Y, \epsilon_1, \epsilon_2, \rho_{CPO})$, by $\widehat{\mathfrak{M}}_{X_{ij}|Y}$, $\widehat{\mathfrak{B}}_{X_iX_j}^y$, $\widehat{\mathfrak{B}}_{XX}^y$, $\widehat{\mathfrak{C}}_{X_iX_j}^y$, $\widehat{\mathfrak{C}}_{XX}^y$, $\widehat{\mathfrak{P}}^y$, and $\widehat{E}_{CPO}$, respectively.

4.2. Empirical Bases and Coordinate Representation

We next introduce a set of empirical bases and the corresponding coordinate representations. We adopt the following notation. Let Ω be a Hilbert space of functions of T, spanned by a set of bases $\mathscr{B} = \{b_1, ..., b_m\}$. For any $\omega \in \Omega$, let $\lfloor \omega \rfloor_{\mathscr{B}} = (\lfloor \omega \rfloor_{\mathscr{B},1}, ..., \lfloor \omega \rfloor_{\mathscr{B},m})^{\top}$ denote its coordinate with respect to $\mathscr{B}$. Then ω can be represented as $\sum_{i-1}^{m} \lfloor \omega \rfloor_{\mathscr{B},i} b_i = \mathscr{B}^{\top} \lfloor \omega \rfloor_{\mathscr{B}}$. Let $K_{\mathscr{B}} = [\langle b_s, b_t \rangle_{\Omega}]_{s,t=1}^{m}$ denote the Gram kernel matrix, which implies that, for any $(\omega_1, \omega_2) \in \Omega$, the inner product $\langle \omega_1, \omega_2 \rangle_{\Omega} = \lfloor \omega_1 \rfloor_{\mathscr{B}}^{\top} K_{\mathscr{B}} \lfloor \omega_2 \rfloor_{\mathscr{B}}$. Throughout the article, we use the symbol $\lfloor \cdot \rfloor$ exclusively for a chosen coordinate system. Let Ω and Ω' be two Hilbert spaces spanned by $\mathscr{B} = \{b_1, ..., b_m\}$ and $\mathscr{B}' = \{b'_1, ..., b'_{m'}\}$, respectively. Let $A : \Omega \rightarrow \Omega'$ be a linear operator. Then the coordinate representation of A with respect to $\mathscr{B}$ and $\mathscr{B}'$ is $\{\lfloor Ab_1 \rfloor_{\mathscr{B}'}, ..., \lfloor Ab_m \rfloor_{\mathscr{B}'}\} \equiv {}_{\mathscr{B}'}\lfloor A \rfloor_{\mathscr{B}}$. For a third Hilbert space Ω'' with base $\mathscr{B}''$, and another linear operator $A' : \Omega' \rightarrow \Omega''$, it is easy to see that ${}_{\mathscr{B}''}\lfloor A'A \rfloor_{\mathscr{B}} = ({}_{\mathscr{B}''}\lfloor A' \rfloor_{\mathscr{B}'})({}_{\mathscr{B}'}\lfloor A \rfloor_{\mathscr{B}})$. For simplicity, we use $\lfloor A \rfloor$ instead of ${}_{\mathscr{B}'}\lfloor A \rfloor_{\mathscr{B}}$ when there is no confusion.

Note that the random function X_i^k can only be observed at a finite set of points. To enable computation, we need to approximate the random functions using the partially observed data. We adopt the construction used in Li and Solea (2018), which assumes the sample path of X^k lies on an RKHS of the time variable T with a finite basis. Specifically, suppose X_i^k is observed on a finite set of time points $T_k = \{T_{k1}, ..., T_{kn_k}\}$, $k = 1, ... n$. Let $\mathscr{T} = (T_{11}, ..., T_{1,n_1}, ..., T_{n1}, ..., T_{nn_n})^{\top} = (\mathscr{T}_1, ..., \mathscr{T}_N)^{\top}$ which pools together all the unique time points across all subjects, and N is the length of $\mathscr{T}$. Letting $\kappa_T : T \times T \rightarrow \mathbb{R}$ be a positive-definite kernel, then Ω_{X_i} can be constructed through $\Omega^N = \text{span}\{\kappa_T(\cdot, \mathscr{T}_1), ..., \kappa_T(\cdot, \mathscr{T}_N)\} = \text{span}\{\mathscr{B}_t^0(\cdot) : t = 1, ..., N\}$, for $i \in V$. Let $K_T = [\kappa_T(\mathscr{T}_s, \mathscr{T}_t)]_{s,t=1}^{N}$ be the $N \times N$ Gram matrix of κ_T, and its eigen-decomposition, $K_T = U_T^1 D_T^1 (U_T^1)^{\top} + U_T^0 D_T^0 (U_T^0)^{\top}$, where $U_T^1 D_T^1 (U_T^1)^{\top}$ is associated with the m leading eigenvalues, and $U_T^0 D_T^0 (U_T^0)^{\top}$ associated with the last $N - m$ eigenvalues. Here we require m N. Therefore, we can construct an orthonormal basis of Ω^N via

$$\mathscr{B}(\cdot) = [\mathscr{B}_1(\cdot), ..., \mathscr{B}_m(\cdot)]^{\top} = (D_T^1)^{-1/2} (U_T^1)^{\top} \mathscr{B}^0(\cdot), \tag{11}$$

where $\mathscr{B}^0(\cdot) = [\mathscr{B}_1^0(\cdot), ..., \mathscr{B}_N^0(\cdot)]^{\top}$. Then X_i^k can be represented as $X_i^k(\cdot) = \sum_{t=1}^{m} \lfloor X_i^k \rfloor_{\mathscr{B},t} \mathscr{B}_t(\cdot) = \mathscr{B}^{\top}(\cdot) \lfloor X_i^k \rfloor_{\mathscr{B}}$. Note that for the kth individual, the function is observed at n_k time points, which implies that $X_i^k(T_k) \equiv [X_i^k(T_{k1}), ..., X_i^k(T_{kn_k})]^{\top} = (\mathscr{B}^k)^{\top} \lfloor X_i^k \rfloor_{\mathscr{B}}$, where $\mathscr{B}^k$ is the $m \times n_k$ matrix $[\mathscr{B}(T_{k1}), ..., \mathscr{B}(T_{kn_k})]$. Therefore, for a given ridge parameter ϵ_T, the coordinate $\lfloor X_i^k \rfloor_{\mathscr{B}}$ can be estimated by

$$\lfloor X_i^k \rfloor = \left[\mathscr{B}^k (\mathscr{B}^k)^{\top} + \epsilon_T I_m \right]^{-1} \mathscr{B}^k [X_i^k(T_k)]. \tag{12}$$

We next derive the coordinate representations of $\widehat{\mathfrak{B}}_{X_i X_i}$, $\mathfrak{B}_{Y X_{ij}}$, $\mathfrak{B}_{YY}$, which then lead to the coordinates of $\widehat{\mathfrak{B}}_{XX}^y$, $\widehat{\mathfrak{C}}_{XX}^y$, and $\widehat{\mathfrak{P}}^y$. Recall $\kappa_Y : \Omega_Y \times \Omega_Y \to \mathbb{R}$ is the kernel used to build the RKHS of Y. Let $K_Y = \left[\kappa_Y\!\left(Y^s, Y^t\right)\right]_{s,t=1}^n$ be the corresponding $n \times n$ Gram matrix of κ_Y, and $\mathcal{H}_Y = \mathrm{span}\left\{\kappa_Y\!\left(\cdot, Y^1\right), \ldots, \kappa_Y\!\left(\cdot, Y^n\right)\right\} \equiv \mathrm{span}\{\mathscr{B}_Y(\cdot)\}$.

Proposition 6.—For $(f, g) \in \Omega^N \times \Omega^N$, $\lfloor \widehat{\mathfrak{B}}_{X_i X_i} \rfloor = E_n\!\left(\lfloor X_i \rfloor \lfloor X_i \rfloor^\top\right)$, $\lfloor \widehat{\mathfrak{B}}_{YY} \rfloor = n^{-1} K_Y$, and

$$\left\lfloor \widehat{\mathfrak{B}}_{Y X_{ij}}(f \otimes g) \right\rfloor = n^{-1}\!\left[\left(\lfloor f \rfloor^\top \lfloor X_i^1 \rfloor \lfloor g \rfloor^\top \lfloor X_j^1 \rfloor\right), \ldots, \left(\lfloor f \rfloor^\top \lfloor X_i^n \rfloor \lfloor g \rfloor^\top \lfloor X_j^n \rfloor\right)\right]^\top.$$

Moreover, for each $i \in V$, let $\left\{\left(\widehat{\lambda}_i^a, \lfloor \widehat{\eta}_i^a \rfloor\right)\right\}_{a=1}^m$ denote the collection of eigenvalue and eigenvector pairs of $\left\lfloor \widehat{\mathfrak{B}}_{X_i X_i} \right\rfloor$; that is $\left\lfloor \widehat{\mathfrak{B}}_{X_i X_i} \right\rfloor \lfloor \widehat{\eta}_i^a \rfloor = \widehat{\lambda}_i^a \lfloor \widehat{\eta}_i^a \rfloor$. Then for each $i \in V$, $k = 1, \ldots, n$, the sample Karhunen-Loève expansion of X_i^k is of the form,

$$X_i^k = \sum_{a=1}^d \widehat{\alpha}_i^a \widehat{\eta}_i^a = \sum_{a=1}^d \left\langle X_i^k, \widehat{\eta}_i^a \right\rangle \widehat{\eta}_i^a = \sum_{a=1}^d \lfloor X_i^k \rfloor^\top \lfloor \widehat{\eta}_i^a \rfloor \widehat{\eta}_i^a. \tag{13}$$

Proposition 7.—Let $\mathscr{B}_i^* = \left\{\widehat{\eta}_i^1, \ldots, \widehat{\eta}_i^d\right\}$, for each $i \in V$. For $y \in \Omega_Y$, the coordinate representation of $\widehat{\mathfrak{B}}_{X_i X_j}^y$ with respect to $\mathscr{B}_i^*$ and $\mathscr{B}_j^*$ is

$$_{\mathscr{B}_i^*}\!\left\lfloor \widehat{\mathfrak{B}}_{X_i X_j}^y \right\rfloor_{\mathscr{B}_j^*} = \left[\left(\mathbf{a}_i^a\right)^\top c(y) \mathbf{a}_j^b\right]_{a,b=1}^d, \tag{14}$$

where $c(y) = \mathrm{diag}[\boldsymbol{\ell}(y)]$ with $\ell(y) = (K_Y + \epsilon_Y I_n)^{-1} \mathscr{B}_Y(y)$, and $\mathbf{a}_i^a = \left(\widehat{\alpha}_i^{1,a}, \ldots, \widehat{\alpha}_i^{n,a}\right)^\top$. Moreover, if $\mathscr{B}^*$ is the collection $\left\{\mathscr{B}_i^* : i \in V\right\}$, then,

$$
\begin{aligned}
_{\mathscr{B}^*}\!\left\lfloor \widehat{\mathfrak{B}}_{XX}^y \right\rfloor_{\mathscr{B}^*} &= \left[_{\mathscr{B}_i^*}\!\left\lfloor \widehat{\mathfrak{B}}_{X_i X_j}^y \right\rfloor_{\mathscr{B}_j^*}\right]_{i,j=1}^p, \\
\left[_{\mathscr{B}^*}\!\left\lfloor \widehat{\mathfrak{C}}_{XX}^y \right\rfloor_{\mathscr{B}^*}\right]_{i,j} &= \begin{cases} \mathfrak{A}_i^{-1/2} \, _{\mathscr{B}_i^*}\!\left\lfloor \widehat{\mathfrak{B}}_{X_i X_j}^y \right\rfloor_{\mathscr{B}_j^*} \mathfrak{A}_j^{-1/2} & \text{when } i \neq j, \\ I_d & \text{when } i = j, \end{cases} \\
_{\mathscr{B}^*}\!\left\lfloor \widehat{\mathfrak{P}}^y \right\rfloor_{\mathscr{B}^*} &= \left(_{\mathscr{B}^*}\!\left\lfloor \widehat{\mathfrak{C}}_{XX}^y \right\rfloor_{\mathscr{B}^*} + \epsilon_2 I_{p \times d}\right)^{-1},
\end{aligned}
\tag{15}
$$

where $\mathfrak{A}_i = {}_{\mathscr{B}_i^*}\!\left\lfloor \widehat{\mathfrak{B}}_{X_i X_i}^y \right\rfloor_{\mathscr{B}_i^*} + \epsilon_1 I_d$, and I_d is the $d \times d$ identity matrix.

Finally, we compute the squared Hilbert–Schmidt norm of $\left[\widehat{\mathfrak{P}}^y\right]_{i,j}$ as

$$
\begin{aligned}
\left\|\left[\widehat{\mathfrak{P}}^y\right]_{i,j}\right\|_{\mathrm{HS}}^2 &= \sum_{a=1}^d \left\langle \left[\widehat{\mathfrak{P}}^y\right]_{i,j} \widehat{\eta}_j^a, \left[\widehat{\mathfrak{P}}^y\right]_{i,j} \widehat{\eta}_j^a \right\rangle \\
&= \sum_{a=1}^d \lfloor \eta_j^a \rfloor_{\mathscr{B}_j^*}^\top \, _{\mathscr{B}_j^*}\!\left[\widehat{\mathfrak{P}}^y\right]_{i,j \, \mathscr{B}_i^*} \, _{\mathscr{B}_i^*}\!\left[\widehat{\mathfrak{P}}^y\right]_{i,j \, \mathscr{B}_j^*} \lfloor \eta_j^a \rfloor_{\mathscr{B}_j^*} \\
&= \left\|_{\mathscr{B}_i^*}\!\left[\widehat{\mathfrak{P}}^y\right]_{i,j \, \mathscr{B}_j^*}\right\|_{\mathrm{F}}^2 = \left\|[_{\mathscr{B}^*}\!\left[\widehat{\mathfrak{P}}^y\right]_{\mathscr{B}^*}]_{i,j}\right\|_{\mathrm{F}}^2,
\end{aligned}
$$

for $(i, j) \in V \times V$ with $i \neq j$, and $\| \cdot \|_F$ is the Frobenius norm.

4.3. Algorithm

We now summarize our conditional functional graph estimation algorithm based on CPO. The algorithm based on CPCO is similar and is thus omitted. Let $\lambda_i(A)$ be the ith largest eigenvalues of an $a \times a$ matrix A, for $i = 1, \ldots, a$.

1. Choose the kernel function κ_T. Some commonly used kernel functions include the Brownian motion function $\kappa_T(s, t) = \min(s, t)$, or the radial basis function (RBF) $\kappa_T(s, t) = \exp[-\gamma_T(s - t)^2]$, $(s, t) \in \mathbb{R}^2$. For RBF, the bandwidth γ_T is determined by $(\sum_{s < t}|\mathcal{T}_s - \mathcal{T}_t|)^2 \gamma_T = N^2(N - 1)^2/4$.

2. Compute the $N \times N$ Gram matrix K_T of κ_T, and let $U_T^1 D_T^1 (U_T^1)$ be its eigen-decomposition associated with its m leading eigenvalues. Then use (11) to construct the reduced basis $\mathscr{B}(\cdot)$, and the matrix $\mathscr{B}^k$. We suggest choosing m by $\min\{m': \sum_{i=1}^{m'}\lambda_i(K_T)/\sum_{i=1}^{N}\lambda_i(K_T) \geq 0.99\}$, in the sense that the cumulative percentage of total variation of K_T explained exceeds 99%.

3. Calculate the coordinate $\lfloor X_k^i \rfloor$, $i = 1, \ldots, p, k = 1, \ldots, n$, using (12) with a given ϵ_T. We choose $\epsilon_T = N^{-1/5}\lambda_1(K_T)$ (see, for example Lee, Li, and Zhao 2016a).

4. Perform eigen-decomposition of $\left\lfloor \widehat{\mathfrak{B}}_{X_i X_i} \right\rfloor$, for $i \in V$, and obtain the ath eigenvector $\lfloor \hat{\eta}_i^a \rfloor$, and the ath K-L coefficient $\hat{\alpha}_i^{k,a}$ using (13), $a = 1, \ldots, d$. Stack $\hat{\alpha}_i^{1,a}, \ldots, \hat{\alpha}_i^{n,a}$ to form $\mathbf{a}_i^a$, $a = 1, \ldots, d, i = 1, \ldots, p, a = 1, \ldots, d, i = 1, \ldots, p$. We choose d according to $\min\left\{d': \sum_{j=1}^{d'}\lambda_j\left(\left\lfloor\widehat{\mathfrak{B}}_{X_i X_i}\right\rfloor\right)/\sum_{j=1}^{N}\lambda_j\left(\left\lfloor\widehat{\mathfrak{B}}_{X_i X_i}\right\rfloor\right) \geq 0.99\right\}$.

5. Choose the kernel function κ_Y, and compute the corresponding $n \times n$ Gram matrix K_Y. Compute the coordinate $_{\mathscr{B}_i^*}\left[\widehat{\mathfrak{B}}_{X_i X_j}^y\right]_{\mathscr{B}_j^*}$ using (14), with the ridge parameter $\epsilon_Y = n^{-1/5}\lambda_1(K_Y)$.

6. Compute the representation of $\widehat{\mathfrak{B}}^y$ using (15), for a given $y \in \Omega_Y$ and the ridge parameters $\epsilon_1 = n^{-1/5}\max\left\{\lambda_1\left(\left\lfloor\widehat{\mathfrak{B}}_{X_i X_i}^y\right\rfloor\right): i \in V\right\}$, and $\epsilon_2 = n^{-1/5}\lambda_1\left(\left\lfloor\widehat{\mathfrak{B}}^y\right\rfloor\right)$.

7. Estimate the conditional functional graph for a given $y \in \Omega_Y$ by $\widehat{E}^y(\rho_{CPO}) = \left\{(i, j) \in V \times V: \left\|\left[\mathscr{B}^*\left[\widehat{\mathfrak{P}}^y\right]\mathscr{B}^*\right]_{i,j}\right\|_F > \rho_{CPO}, i \neq j\right\}$, with a given threshold ρ_{CPO}. We determine ρ_{CPO} by minimizing the following generalized cross-validation (GCV) criterion over a set of grid points,

$$\text{GCV}^y(\rho) = \sum_{i=1}^{p}\text{GCV}_i^y(\rho), \text{ with}$$

$$\text{GCV}_i^y(\rho) = \frac{\left\|_{\mathscr{B}_i^*}\left[\widehat{\mathfrak{B}}_{X_i X_i \mid X_{N_i^y(\rho)}}^y\right]_{\mathscr{B}_i^*}\right\|_F^2/n}{[1 - \text{DF}_i(\rho)/n]^2},$$

where $\widehat{\mathfrak{B}}^y_{X_iX_i \mid X_{N^y_i(\rho)}}$ is the sample estimate of $\widehat{\mathfrak{B}}^y_{X_iX_i \mid X_{N^y_i(\rho)}}$, and its coordinate representation is derived in Section S.4 of the appendix, and $\mathrm{DF}_i(\rho) = d^2$ card $[\mathrm{N}_i(\rho)]$, with card$(\cdot)$ being the cardinality and $\mathrm{N}^y_i(\rho) = \left\{ j \in \mathrm{V} : (i,j) \in \widehat{\mathrm{E}}^y(\rho) \right\}$ the neighborhood of the ith node in $\widehat{\mathrm{E}}^y(\rho)$.

Our graph estimation algorithm involves multiple tuning parameters, and their choices are given in the above algorithm. We further study their effects in Section S.6 of the appendix. In general, we have found that the estimated graph is not overly sensitive to the tuning parameters as long as they are within a reasonable range.

We also remark that, the above algorithm assumes the zero-mean condition. In practice, if this does not hold, we can easily modify the algorithm. Specifically, from the coordinates of the estimated conditional covariance operator, we can estimate $E(\alpha^a_i \mid y)$ by $(a^a_i)^\top c(y) 1_n$, where 1_n is the n-dimensional vector with all ones. By redefining the conditional covariance operator as $\mathfrak{B}^y_{X_iX_j} = \sum_{a,b \in \mathbb{N}} [E(\alpha^a_i \alpha^b_j \mid y) - E(\alpha^a_i \mid y) E(\alpha^b_j \mid y)](\eta^a_i \otimes \eta^b_j)$, we can estimate the coordinates of $\widehat{\mathfrak{B}}^y_{X_iX_j}$ by $_{\mathscr{B}^*_i} \left[\widehat{\mathfrak{B}}^y_{X_iX_j} \right]'_{\mathscr{B}^*_j} = \left[(\mathbf{a}^a_i)^\top c'(y) \mathbf{a}^b_j \right]^d_{a,b=1}$, where $c'(y) = \mathrm{diag}[\ell(y)] - \ell(y)$ $\ell(y)^\top$. We then replace $_{\mathscr{B}^*_i} \left[\widehat{\mathfrak{B}}^y_{X_iX_j} \right]_{\mathscr{B}^*_j}$ in Step 5 of the algorithm, and all subsequent procedures, by $_{\mathscr{B}^*_i} \left[\widehat{\mathfrak{B}}^y_{X_iX_j} \right]_{\mathscr{B}^*_j}$.

5. Asymptotic Theory

We begin with some useful concentration inequalities and uniform convergence for several relevant operators. We next establish the uniform convergence of the CPO, and the consistency of the estimated graph. For simplicity, we first assume that the trajectory of the random functions $X(t) = [X_1(t), \ldots, X_p(t)]^\top$ is fully observed for all $t \in T$. We then discuss the setting when X_i is only partially observed in Section 5.3. We also remark that, all our theoretical results allow the dimension of the graph to diverge with the sample size.

5.1. Concentration Inequalities and Uniform Convergence

We first derive the concentration bound and uniform convergence rate for the sample estimators $\widehat{\mathfrak{B}}_{X_iX_j}$, $\widehat{\mathfrak{B}}_{YX_{ij}}$, and $\widehat{\mathfrak{B}}_{YY}$. For two positive sequences $\{a_n\}$ and $\{b_n\}$, write $a_n \prec b_n$ if $a_n = o(b_n)$, $a_n \precsim b_n$ if $a_n = O(b_n)$, and $a_n \asymp b_n$ if $a_n \precsim b_n$ and $b_n \precsim a_n$; moreover, let $a_n \vee b_n = b_n$ if $a_n \precsim b_n$. Similarly, if c_n is a third positive sequence, then we let $a_n \vee b_n \vee c_n = (a_n \vee b_n) \vee c_n$.

Theorem 3.—If Assumptions 1 and 3 hold, then there exist positive constants C_1 to C_6, such that, (i) $P\left(\left\| \widehat{\mathfrak{B}}_{X_iX_j} - \mathfrak{B}_{X_iX_j} \right\|_{\mathrm{HS}} > t \right) \leq C_1 \exp\left[-C_2 n \left(t \wedge t^2 \right) \right]$; (ii) $P\left(\left\| \widehat{\mathfrak{B}}_{YX_{ij}} - \mathfrak{B}_{YX_{ij}} \right\|_{\mathrm{HS}} > t \right) \leq C_3 \exp\left[-C_4 n \left(t \wedge t^2 \right) \right]$; (iii) $P\left(\left\| \widehat{\mathfrak{B}}_{YY} - \mathfrak{B}_{YY} \right\|_{\mathrm{HS}} > t \right) \leq C_5 \exp\left(-C_6 n t^2 \right)$, for any $t \ge 0$ and any $(i,j) \in \mathrm{V} \times \mathrm{V}$. Moreover, if $\log p / n \to 0$, then (iv) $\max_{i,j \in \mathrm{V}} \left\| \widehat{\mathfrak{B}}_{X_iX_j} - \mathfrak{B}_{X_iX_j} \right\|_{\mathrm{HS}} = O_P\left[(\log p / n)^{1/2} \right]$; (v) $\max_{i,j \in \mathrm{V}} \left\| \widehat{\mathfrak{B}}_{YX_{ij}} - \mathfrak{B}_{YX_{ij}} \right\|_{\mathrm{HS}} = \left\| \mathfrak{B}_{X_iX_j} \right\|_{\mathrm{HS}} = O_P\left[(\log p / n)^{1/2} \right]$; (vi) $\left\| \widehat{\mathfrak{B}}_{YY} - \mathfrak{B}_{YY} \right\|_{\mathrm{HS}} = O_P\left(n^{-1/2} \right)$, as $n \to \infty$.

For any $(i, j) \in V \times V$ and $y \in \Omega$, let $\mathfrak{B}^y_{X_i X_j}(d)$ be the truncated version of $\mathfrak{B}^y_{X_i X_j}$: $\mathfrak{B}^y_{X_i X_j}(d) = \sum_{a,b=1}^{d} E(\alpha_i^a \alpha_j^b \mid y)(\eta_i^a \otimes \eta_j^b)$. Next, we establish the uniform convergence rate for $\left\| \widehat{\mathfrak{B}}^y_{X_i X_j} - \mathfrak{B}^y_{X_i X_j}(d) \right\|_{HS}$. We define the exponent of a compact and self-adjoint operator as $A^\beta = \sum_{a \in \mathbb{N}} \lambda_a^\beta (\eta_a \otimes \eta_a)$, for any $\beta > 0$, where $\{(\lambda_a, \eta_a)\}_{a \in \mathbb{N}}$ is the collection of eigenvalue and eigenfunction pairs of A. We need another assumption.

Assumption 6.—For any $(i, j) \in V \times V$, there exist $\beta \in (0, 1)$, $c_2 > 0$, and $\mathfrak{M}^0_{ij} \in \mathscr{B}[\mathscr{B}_2(\Omega_j, \Omega_i), \mathscr{H}_Y]$ such that $\mathfrak{M}_{X_{ij} \mid Y} = \mathfrak{B}^\beta_{YY} \mathfrak{M}^0_{ij}$ with $\|\mathfrak{M}^0_{ij}\| \le c_2$.

This assumption regulates the complexity of $(X_i, X_j)^\top$ given Y. To see this, note that, as β increases, the regression relation from $(X_i, X_j)^\top$ on Y are more concentrated on those eigenfunctions corresponding to larger eigenvalues of $\mathfrak{B}_{YY}$. Also, we define $\kappa_d \equiv \min\{|\lambda_i^a - \lambda_i^b| : 1 \le a < b \le d + 1, i \in V\}$ as the minimal isolation distance among all $d + 1$ leading eigenvalues of $\mathfrak{B}_{X_i X_i}$, for all $i \in V$. Note that we allow d to grow with n.

Theorem 4.—If Assumptions 1–3 and 6 hold, $\epsilon_Y \prec 1$, and $(\log p / n) \prec \kappa_d^2$, then
$$\max_{i,j \in V} \left\| \widehat{\mathfrak{B}}^y_{X_i X_j} - \mathfrak{B}^y_{X_i X_j}(d) \right\|_{HS} = O_P \left[d^2 \epsilon_Y^{-1} \kappa_d^{-1} (\log p / n)^{1/2} + d^2 \epsilon_Y^{-1} n^{-1/2} + d^2 \epsilon_Y^\beta \right].$$

Next, we establish the uniform convergence of the estimated conditional correlation operator $\widehat{\mathfrak{C}}^y_{XX}$. We need an assumption on the tail behavior of random functions.

Assumption 7.—There exists $\gamma_y > 0$ such that $\max_{i \in V} \left\{ \sum_{a=d+1}^{\infty} E\left[(\alpha_i^a)^2 \mid y \right] \right\} \le d^{-\gamma_y}$, for any $y \in \Omega_Y$.

Assumption 7 is on the decaying rate of the tail eigenvalues of $\mathfrak{B}^y_{X_i X_i}$, which, in a sense, characterizes the smoothness of the distribution of X_i given Y.

Theorem 5.—If Assumptions 1–4, 6, and 7 hold, ϵ_Y, $\epsilon_1 \prec 1$, and $d^2 \epsilon_Y^{-1} \kappa_d^{-1} (\log p / n)^{1/2} + d^2 \epsilon_Y^{-1} n^{-1/2} + d^2 \epsilon_Y^\beta \prec 1$, then, for any $y \in \Omega_Y$,

$$\max_{i,j \in V} \left\| \left[\widehat{\mathfrak{C}}^y_{XX} \right]_{i,j} - \left[\mathfrak{C}^y_{XX} \right]_{i,j} \right\|_{HS} = O_p(\delta_y),$$

where $\delta_y = \epsilon_1^{-3/2} \left[d^2 \epsilon_Y^{-1} \kappa_d^{-1} (\log p / n)^{1/2} + d^2 \epsilon_Y^{-1} n^{-1/2} + d^2 \epsilon_Y^\beta + d^{-\gamma_y} \right] + \epsilon_1^{1/2}$.

To better understand Theorem 5, suppose $d = O(n^{a_1/2})$, $\epsilon_Y = O(n^{-a_2})$, $\epsilon_1 = O(n^{-2a_3})$, and $k_d = O(n^{-a_4})$, for $a_1, \ldots, a_4 > 0$. Then we have

$$\max_{i,j \in V} \left\| \left[\widehat{\mathfrak{C}}^y_{XX} \right]_{i,j} - \left[\mathfrak{C}^y_{XX} \right]_{i,j} \right\|_{HS} = O_p \left[\left(\log p / n^{1-\pi} \right)^{1/2} + n^{-\frac{1}{2} a_1 \gamma_y + 3a_3} + n^{-a_2 \beta + a_1 + 3a_3} + n^{-\frac{1}{2} + a_1 + a_2 + 3a_3} + n^{-a_3} \right],$$

where $\pi = 2(a_1 + a_2 + a_4) + 6a_3 < 1$. This implies that the graph dimension p can diverge with the sample size n at an exponential rate. In comparison, in the classical random variable setting, the uniform convergence rate of the sample covariance is $(\log p/n)^{1/2}$ (Bickel and Levina 2008). Theorem 5 thus can be viewed as an extension to both the functional and conditional settings where the parameter of interest $\mathfrak{C}_{XX}^y$ is a random RKHS operator.

5.2. Uniform Convergence of CPO and Graph Consistency

We next derive the convergence of the estimated CPO, $\widehat{\mathfrak{P}}^y$. We need an assumption to regulate the relation between X_i and X_j when conditioning on $X_{-(i,j)}$ and Y.

Assumption 8.—There exists $c_3 > 0$ such that, $\max_{i,j \in \mathrm{V}, i \neq j} \left\| \mathfrak{C}_{X_i X_j \mid X_{-(i,j)}}^y \right\|_{\mathrm{HS}} \leq c_3$, where

$$\mathfrak{C}_{X_i X_j \mid X_{-(i,j)}}^y = \mathfrak{C}_{X_i X_j}^y - \mathfrak{C}_{X_i X_{-(i,j)}}^y \left[\mathfrak{C}_{X_{-(i,j)} X_{-(i,j)}}^y \right]^{-1} \mathfrak{C}_{X_{-(i,j)} X_j}^y, \text{ for any } y \in \Omega_Y$$

The following proposition provides a condition under which Assumption 8 is satisfied. Its proof is similar to Proposition S2 and is omitted.

Proposition 8.—Suppose Assumptions 1, 3–5 hold. Then, for any $y \in \Omega_Y$, there exists $c_3 > 0$ such that, $\max_{i,j \in \mathrm{V}, i \neq j} \| \mathfrak{C}_{X_i X_j \mid X_{-(i,j)}}^y \|_{\mathrm{HS}} \leq c_3$, if there exists $c4 > 0$ satisfying that

$$\max_{i, j \in \mathrm{V}, i \neq j} \sum_{a \in \widetilde{\mathbb{N}}_i^y, b \in \widetilde{\mathbb{N}}_j^y} \mathrm{cor}^2 \left(\langle \widetilde{\psi}_i^{y,a}, X_i \rangle, \langle \widetilde{\psi}_j^{y,b}, X_j \rangle \mid X_{-(i,j)}, y \right) \leq c_4, \tag{16}$$

where $\widetilde{\psi}_i^{y,a} = \psi_i^{y,a}/(\mu_i^{y,a})^{1/2}$, and $\{(\mu_i^{y,a}, \psi_i^{y,a})\}_{a \in \mathbb{N}}$ are the pairs of eigenvalue and eigenfunction of $\mathfrak{B}_{X_i X_i \mid X_{-(i,j)}}^y$, and $\widetilde{\mathbb{N}}_i^y = \{a \in \mathbb{N} : \mu_i^{y,a} > 0\}$, for any $i \in \mathrm{V}$ and $y \in \Omega_Y$.

Both conditions (5) and (16) introduce certain levels of smoothness on the conditional distribution of X given Y. Nevertheless, they target different subjects: Equation (5) is about the relation between X_i and X_j given Y, which is required for the consistency of $\mathfrak{C}_{X_i X_j'}^y$, whereas Equation (16) is about the relation between X_i and X_j given $X_{-(i,j)}$ and Y, which is required for the consistency of $[\mathfrak{P}^y]_{i,j}$.

Theorem 6.—If Assumptions 1–8 hold, $\epsilon_Y, \epsilon_1 \prec 1$, and $d^2 \epsilon_Y^{-1} \kappa_d^{-1} (\log p/n)^{1/2} + d^2 \epsilon_Y^{-1} n^{-1/2} + d^2 \epsilon_Y^\beta \prec 1$, then for any $y \in \Omega_Y$,

$$\max_{i, j \in \mathrm{V}, i \neq j} \left\| [\widehat{\mathfrak{P}}^y]_{i,j} - [\mathfrak{P}^y]_{i,j} \right\|_{\mathrm{HS}} = O_p \left(\epsilon_2^{-1} p \delta_y + \epsilon_2 \right).$$

Moreover, if $\rho_{\mathrm{CPO}} = \min \{ \| [\mathfrak{P}^y]_{i,j} \|_{\mathrm{HS}} : (i, j) \in \mathrm{V} \times \mathrm{V}, i \neq j, [\mathfrak{P}^y]_{i,j} \neq 0 \}/2$, and $\rho_{\mathrm{CPO}} \succ (\epsilon_2^{-1} p \delta_y + \epsilon_2)$, then,

$$P \left[\widehat{\mathrm{E}}_{\mathrm{CPO}}^y = \mathrm{E}^y \right] \to 1, \text{ as } n \to \infty.$$

We make a few remarks. First, Li and Solea (2018) established the consistency of their precision operator, as well as the unconditional graph estimation consistency. Note that their operator is nonrandom, and their results were derived with the graph size p fixed. By contrast, Theorem 6 establishes the uniform convergence of the operator $\left[\widehat{\mathfrak{P}}^y\right]_{i,j}$, which is random, and the graph estimation consistency is obtained with a diverging p. For instance, if $\epsilon_2 \ O(n^{-\pi'/2})$ with $\pi' > 0$, then Theorem 6 says that the uniform convergence rate of the estimated CPO depends on $p(\log p/n^{1-\pi-\pi'})^{1/2}$. This implies that we allow p to grow at the polynomial rate of n.

Second, a careful inspection of our proof reveals that, the convergence rate of the empirical CPO depends on the difference $\|\widehat{\mathfrak{C}}^y_{XX} - \mathfrak{C}^y_{XX}\|_{\mathrm{HS}}$, whose order, by Theorem 5, can be no faster than $p(\log p/n^{1-\pi'})^{1/2}$. We note that, under the classical random variable setting, the convergence rate of the hard thresholding sample covariance matrix in terms of Frobenius norm is $[pc_0(p) \log p/n]^{1/2}$, where the term $c_0(p)$ imposes a sparsity structure on the covariance matrix and satisfies that $\sum_{j=1}^p \mathbf{1}[\mathrm{cov}(X_i, X_j) \neq 0] \leq c_0(p)$, for $i = 1, \ldots, p$, with $\mathbf{1}(\cdot)$ being the indicator function (Bickel and Levina 2008, theo. 2). We feel our rate of $p(\log p/n^{1-\pi})^{1/2}$ is reasonable and is comparable to the classical result. We also note that there is a difference between p and $p^{1/2}$ in our rate and the rate of (Bickel and Levina 2008, theo. 2). This difference is mainly due to different sparsity settings imposed by our method and by Bickel and Levina (2008). Note that we do *not* impose any sparsity structure on the conditional correlation operator $\mathfrak{C}^y_{XX}$. This means we need to estimate all the off-diagonal elements of $\mathfrak{C}^y_{XX}$, whose cardinality grows in the order of p^2. In comparison, Bickel and Levina (2008) imposed a sparsity structure on the covariance matrix and the cardinality of nonzero covariances grows only in the order of $pc_0(p)$. Moreover, we are dealing with a more complicated setting of random functions and random operators. On the other hand, we show in Section S.5 in the appendix that, if we introduce additional sparsity and regularization, we can further improve the rates in Lemma S8 in the appendix and Theorem 6, so that p can grow at an exponential rate of n. Actually, we have developed a theoretical platform that is not only limited to the present setting. The concept of using vanishing CPO to identify the conditional independence between random functions in a conditional graph can have divarication beyond the scenarios studied in this article; for example, when there are additional sparsity or bendable structures.

Finally, in our theory development, we have imposed a series of technical conditions, which are commonly imposed in the literature, and are usually easy to satisfy.

5.3. Consistency for Partially Observed Random Functions

We next derive the consistency under the scenario when the random functions are only partially observed. Partially observed functions are collected via a dense or a non-dense measurement schedule; see Wang, Chiou, and Muller (2016a) for more discussion on measurement schedule. To avoid digressing from the main context, in this article, we do not go after any specific measurement schedule or regularity setting on the partially observed random function. For partially observed functions $X(t)$, suppose $\widetilde{X}(t) = \left[\widetilde{X}_1(t), \ldots, \widetilde{X}_p(t)\right]^\top$ is

the estimate of $X(t)$ using the empirical bases developed in Section 4.2. We then estimate the series of the operators and the graph by

$$\widetilde{\mathfrak{B}}_{X_iX_i} = E_n(\widetilde{X}_i \otimes \widetilde{X}_i), \quad \widetilde{\mathfrak{B}}_{YX_{ij}} = E_n[\kappa_Y(\,\cdot\,, Y) \otimes \widetilde{X}_i \otimes \widetilde{X}_j],$$

$$\widetilde{\mathfrak{B}}^y_{X_iX_j} = \sum_{a,b=1}^{d} \left\{\left[\widetilde{\mathfrak{M}}_{X_{ij}\mid Y}(\tilde{\eta}_i^a \otimes \tilde{\eta}_j^b)\right]^\circ(y)\right\}(\tilde{\eta}_i^a \otimes \tilde{\eta}_j^b),$$

$$\widetilde{\mathfrak{C}}^y_{X_iX_j} = \left(\widetilde{\mathfrak{B}}^y_{X_iX_i} + \epsilon_1 I\right)^{-1/2}\widetilde{\mathfrak{B}}^y_{X_iX_j}\left(\mathfrak{B}^y_{X_jX_j} + \epsilon_1 I\right)^{-1/2}$$
for $i \neq j$, and I for $i = j$,

$$\widetilde{\mathfrak{P}}^y = \left(\widetilde{\mathfrak{C}}^y_{XX} + \epsilon_2 I\right)^{-1},$$

$$\widetilde{\mathrm{E}}^y(\rho) = \left\{(i,j) \in \mathrm{V} \times \mathrm{V} : \left\|\left[\widetilde{\mathfrak{P}}^y\right]_{i,j}\right\|_{\mathrm{HS}} > \rho, i \neq j\right\},$$

where $\widetilde{\mathfrak{M}}_{X_{ij}\mid Y} = \left(\widehat{\mathfrak{B}}_{YY} + \epsilon_Y I\right)^{-1}\widetilde{\mathfrak{B}}_{YX_{ij}}$, $\tilde{\eta}_i^a$ is the eigenfunction of $\widetilde{\mathfrak{B}}_{X_iX_i}$, and $\widetilde{\mathfrak{C}}^y_{XX}$ is the block matrix of operators with (i, j)th element being $\widetilde{\mathfrak{C}}^y_{X_iX_j}$.

Theorems 5 and 6 show that the convergence of the estimated CCO and CPO depends on the uniform convergence of the sample covariance operators in Theorem 3. In particular, it relies on the convergence rates of $\max_{i,j \in \mathrm{v}}\|\widehat{\mathfrak{B}}_{X_iX_j} - \mathfrak{B}_{X_iX_j}\|_{\mathrm{HS}}$ and $\max_{i,j \in \mathrm{v}}\|\widehat{\mathfrak{B}}_{YX_{ij}} - \mathfrak{B}_{YX_{ij}}\|_{\mathrm{HS}}$. When the random functions are completely observed, both rates, by Theorem 3, are equal to $(\log p/n)^{1/2}$. When the random functions are only partially observed, we let the convergence rates of $\max_{i,j \in \mathrm{v}}\|\widehat{\mathfrak{B}}_{X_iX_i} - \mathfrak{B}_{X_iX_i}\|_{\mathrm{HS}}$ and $\max_{i,j \in \mathrm{v}}\|\widehat{\mathfrak{B}}_{YX_{ij}} - \mathfrak{B}_{YX_{ij}}\|_{\mathrm{HS}}$ to be slower than $(\log p/n)^{1/2}$. More specifically, as specified in Equation (17), we introduce a quantity a that reflects how dense the time points are observed in the random functions. The denser the time points are, the closer a is to zero. Correspondingly, the next theorem extends Theorem 6 to partially observed functions. Its proof is similar to that of Theorem 6, and is thus omitted.

Theorem 7.—If Assumptions 1–8 hold, ϵ_Y, $\epsilon_1 \prec 1$, and there exists $a \in [0, 1)$ such that $d^2\epsilon_Y^{-1}\kappa_d^{-1}\log p/n^{(1-a)/2} + d^2\epsilon_Y^{-1}n^{-1/2} + d^2\epsilon_Y^\beta \prec 1$, $[\epsilon_2^{-1}p\delta_y^a + \epsilon_2] \prec \rho_{\mathrm{CPO}}$ with $\delta_y^a = \epsilon_1^{-3/2}\left\{d^2\epsilon_Y^{-1}\kappa_d^{-1}\log p/n^{(1-a)/2} + d^2\epsilon_Y^{-1}n^{-1/2} + d^2\epsilon_Y^\beta + d^{-\gamma_y}\right\} + \epsilon_1^{1/2}$, and

$$\begin{aligned}
\max_{i,j \in \mathrm{V}} \|\widetilde{\mathfrak{B}}_{X_iX_j} - \mathfrak{B}_{X_iX_j}\|_{\mathrm{HS}} &= O_P\left[\log p/n^{(1-a)/2}\right], \\
\max_{i,j \in \mathrm{V}} \|\widetilde{\mathfrak{B}}_{YX_{ij}} - \mathfrak{B}_{YX_{ij}}\|_{\mathrm{HS}} &= O_P\left[\log p/n^{(1-a)/2}\right].
\end{aligned} \tag{17}$$

Then, $\max\left\{\left\|\left[\widetilde{\mathfrak{P}}^y\right]_{i,j} - \left[\mathfrak{P}^y\right]_{i,j}\right\|_{\mathrm{HS}} : (i, j) \in \mathrm{V} \times \mathrm{V}, i \neq j\right\} = O_p\left[\epsilon_2^{-1} p \delta_y^a + \epsilon_2\right]$, and $P\left[\widetilde{\mathrm{E}}^y(\rho_{\mathrm{CPO}}) = \mathrm{E}^y\right] \rightarrow 1$,

for any $y \in \Omega_Y$, as $n \rightarrow \infty$.

The first condition of Equation (17) is satisfied if the tail of $\|\widetilde{\mathfrak{B}}_{X_i X_j} - \mathfrak{B}_{X_i X_j}\|_{\mathrm{HS}}$ behaves as a sub-Exponential distribution. That is, when there exist c', $c'' > 0$ and $a \in [0, 1)$, such that $P\left(\|\widetilde{\mathfrak{B}}_{X_i X_j} - \mathfrak{B}_{X_i X_j}\|_{\mathrm{HS}} > t\right) \leq c' \exp\left(-c_2'' n^{(1-a)/2} t\right)$, for any $t \geq 0$. A similar condition also holds for the second condition of Equation (17). Recall in Theorem 3, we have shown that, for completely observed functions, $\|\widehat{\mathfrak{B}}_{X_i X_j} - \mathfrak{B}_{X_i X_j}\|_{\mathrm{HS}} > t$ and $\|\widehat{\mathfrak{B}}_{Y X_{ij}} - \mathfrak{B}_{Y X_{ij}}\|_{\mathrm{HS}} > t$ behave as a sub-Gaussian distribution for a small t. As such, Equation (17) is more appropriate for partially observed functions.

6. Simulations

Next, we study the finite-sample performance of our method through simulations. Specifically, we consider three graph structures of E: a hub, a tree, and a random graph, as shown in Table 1. We consider $p = 10$ nodes with the sample size $n = 100$ first, and consider larger graphs later. Given the graph structure E and the conditional variable y, we generate $X_j(t)$ and its parent nodes based on the following model:

$$X_j(t) \mid y = \sum_{i \in P_j}\left[\sqrt{y}^{\upsilon_{ij}}(1 - \sqrt{y})^{1 - \upsilon_{ij}}\right] \times X_i(t) + c_j \times \varepsilon_j(t),$$

where $X(t) \mid y = [X_1(t) \mid y, \ldots, X_p(t) \mid y]^\top$ is constructed sequentially via the given graph, P_j is the set of the parent nodes of j, and c_j is the scale parameter as specified in Table 1. We generate the error function, $\varepsilon_j(t)$ as $\sum_{u=1}^{J} \xi_u \kappa_T(t, t_u)$, where $\xi_1, \ldots, \xi_J$ are iid standard normal variables, $t_1, \ldots, t_J$ are equally spaced points between $[0, 1]$ with $J = 50$, and κ_T is a RBF or a Brownian motion covariance function. We then generate the conditioning variable, $Y^1, \ldots, Y^n$, as iid Uniform$(0, 1)$. We generate each X_i^k, $k = 1, \ldots, n$, $i = 1, \ldots, p$, with $n_k = 50$ time points. In this model, there are two types of edge patterns: when $\upsilon_{ij} = 1$, the strength of edges grows with y, and when $\upsilon_{ij} = 0$, the strength of edges decays with y. The tuning parameters are determined by the rules suggested in Section 4.3. In Section S.6 of the appendix, we discuss in detail the effect of the tuning parameters, and show that our method is relatively robust to a range of tuning parameters.

We compare our method with three alternative solutions, all of which are variants of graphical Lasso (Friedman et al. 2008, gLASSO). The first solution, which we refer as "Average," is similar to Kolar et al. (2010). It first estimates the conditional covariance matrix by $\widehat{\Sigma}_{XX}^y = \sum_{k=1}^{n} \kappa_y(y, Y^k)\left\{\sum_{j=1}^{50} X^k(T_j)\left[X^k(T_j)\right]^\top / 50\right\} / \sum_{k=1}^{n} \kappa_y(y, Y^k)$, where $X^k(T_j) = \left[X_1^k(T_j), \ldots, X_p^k(T_j)\right]^\top$, $j = 1, \ldots, 50$, $k = 1, \ldots, n$. It then estimates the conditional precision matrix $\widehat{\theta}^y$ by applying gLASSO to $\widehat{\Sigma}_{XX}^y$. The second solution, which we refer as "Majority," is similar to a procedure in Qiao, Guo, and James (2019). It first estimates the covariance matrix at each time point by $\widehat{\Sigma}_{XX}^y(T_j) = \sum_{k=1}^{n} \kappa_y(y, Y^k)\left\{X^k(T_j)\left[X^k(T_j)\right]^\top\right\} / \sum_{k=1}^{n} \kappa_y(y, Y^k)$. It then estimates the conditional

precision matrix at each time point, by applying gLASSO to $\widehat{\Sigma}_{XX}^y(T_j)$. It selects those edges that are detected by the majority of the estimates among all the estimated graphs. The third solution, referred as "Unconditional," is similar to the naive procedure in Qiu et al. (2016), which, without using the information of Y, estimates the covariance matrix by $\widehat{\Sigma}_{XX}^k = \sum_{j=1}^{50} X^k(T_j)\left[X^k(T_j)\right]^\top/50$. It then estimates the conditional precision matrix by applying gLASSO to $\widehat{\Sigma}_{XX}^k$. For the penalty parameter in gLASSO, we adopt the empirical rule suggested by Rothman et al. (2008), and set it to $(\log p/n)^{1/2}$. We use the RBF kernel for both κ_T and κ_Y in all simulations.

We evaluate the accuracy of the estimated graph using the area under ROC curve (AUC). We first compute the false positive (FP) and true positive (TP) as,

$$\mathrm{TP}^{y,\rho} = \frac{\sum_{1 \le i < j \le p} \mathbf{1}\left[(i,j) \in \mathrm{E}^0, (i,j) \in \widehat{\mathrm{E}}^y(\rho)\right]}{\sum_{1 \le i < j \le p} \mathbf{1}\left[(i,j) \in \mathrm{E}^0\right]},$$

$$\mathrm{FP}^{y,\rho} = \frac{\sum_{1 \le i < j \le p} \mathbf{1}\left[(i,j) \notin \mathrm{E}^0, (i,j) \in \widehat{\mathrm{E}}^y(\rho)\right]}{\sum_{1 \le i < j \le p} \mathbf{1}\left[(i,j) \notin \mathrm{E}^0\right]},$$

for a given ρ and $y \in \Omega_Y$, and E^0 is the true graph. We then compute the pairs of ($\mathrm{TP}^{y,\rho}$, $\mathrm{FP}^{y,\rho}$) over a set of values of ρ, which we set to be the sorted norms of empirical CPO.

Figure 1 reports the AUC for the estimated graph with respect to the external variable Y, under three graph structures, and by the four methods. It is seen that the methods "Majority" and "Unconditional" perform consistently the worst, while "Average" and our proposed CPO methods perform similarly in this example.

We next extend our simulations to larger graphs, where we set the graph size $p = \{30, 50, 100\}$, with the sample size $n = 30$ and $n_k = 30$ time points, $k = 1, \ldots, n$. We then generate the graph in a similar way as before. Specifically, for the hub structure, we generate $p/5$ independent hubs, and within each hub, two edges have their strength growing with y, and two edges decaying with y. For the tree structure, we expand the tree until the graph reaches the designated size, and in each subsequent layer of the tree, one edge has growing strength and the other has decaying strength with y. For the random structure, we select the edges randomly following a Bernoulli distribution with probability $1/(p-1)$, and also randomly select half of the selected edges to have growing strength and the other half to have decaying strength. Due to the poor performance of "Majority" and "Unconditional" in the previous simulation setting, we only compare our CPO method with "Average" in the large-graph setting. Figure 2 reports the AUC. It is seen that our CPO method clearly outperforms the "Average" method in some settings, and is comparable in other settings. In particular, for a large graph with $p = 100$, the improvements of CPO over "Average" in both the hub and random structures are substantial.

7. Application

In this section, we illustrate our conditional functional graph estimation method with a brain functional connectivity analysis example. We analyze a dataset from the Human Connectome Project (HCP), which consists of resting-state fMRI scans of 549 individuals. Each fMRI scan has been processed and summarized as a spatial temporal matrix, with rows corresponding to 268 brain regions-of-interest, and columns corresponding to 1200 time points (Greene et al. 2018). Additionally, each subject is collected with a score of the Penn Progressive Matrices, which is a nonverbal group test typically used in educational settings and is generally taken as a surrogate of fluid intelligence. It is of scientific interest to unravel the connectivity patterns of the brain regions conditioning on the intelligence measure.

We apply our conditional functional graphical modeling approach for the whole brain of 268×268 connectivity network. Applying the hard thresholding approach would yield a sparse estimate of the conditional graph at each value of $Y = y$. To identify edges that vary along with the conditional variable Y, we further employ a permutation test approach. Specifically, we permute the observations of Y five hundred times. For each permuted sample, we compute the Hilbert-Schmidt (H-S) norm of the sample CPO for each edge. We then compute the variance of the H-S norms based on 500 permutations. We treat those edges whose corresponding variances above 95% percentile as significant, since in this context a nonzero variance implies that the CPO varies with the conditional variable. In addition, those 268 brain regions have been partitioned into eight functional modules: medial frontal, frontoparietal, default mode, motor, visual, limbic, basal ganglia, and cerebellum (Finn et al. 2015). We count the number of significant edges from the permutation test within each functional module, then use the Fisher's exact test to determine if a module is significantly enriched.

We have found that the medial frontal module is significantly enriched with numerous significant edges that vary along with the intelligence score. Figure 3 shows the identified significant edges. This finding agrees with the literature, as this module has been reported to contribute to high-level cognitive functions such as emotional responses (Smith et al. 2018) and learning (Zanto and Gazzaley 2013). Moreover, this module was found to have more impact on fluid intelligence compared to other modules (Finn et al. 2015). We have also observed more cross-lobe and inter-hemisphere interactions, which suggests the importance of these edges in cognitive functions. This again complies with the existing literature in neuroscience that the altered interhemispheric interactions of prefrontal lobe is closely related to higher level cognitive traits such as the Internet gaming disorder (Wang et al. 2015). Figure 4 reports the changes of the identified significant edges for the medial frontal module at five different values of the intelligence score. We see from the plot that both increasing and decreasing patterns exist for the strength of the significantly varying edges.

As an independent validation, we replicate the analysis using another resting-state fMRI data of 828 individuals in HCP. We report the changes of the identified significant edges for the medial frontal module from the new dataset in Figure S6 of the appendix. We again find that the medial frontal module is enriched. Hearne, Mattingley, and Cocchi (2016) identified positive correlations between functional connectivity and intelligence in general.

Song et al. (2008) reported that most of the brain regions whose connectivity patterns are negatively correlated with the intelligence are around medial frontal gyrus, or part of motor regions, which are part of medial frontal module. Combined with our findings, it suggests that the medial frontal module may play a unique role in intelligence compared to other brain modules.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

Bach F (2009), "High-Dimensional Non-Linear Variable Selection Through Hierarchical Kernel Learning," arXiv no. 0909.0844.

Baker CR (1973), "Joint Measures and Cross-Covariance Operators," Transactions of the American Mathematical Society, 186, 273–289.

Bickel PJ, and Levina E (2008), "Covariance Regularization by Thresholding," The Annals of Statistics, 36, 2577–2604.

Bosq D (2000), Linear Processes in Function Spaces, New York: Springer.

Bullmore E, and Sporns O (2009), "Complex Brain Networks: Graph Theoretical Analysis of Structural and Functional Systems," Nature Reviews. Neuroscience, 10, 186–198. [PubMed: 19190637]

Cai TT, Liu W, and Luo X (2011), "A Constrained ℓ_1 Minimization Approach to Sparse Precision Matrix Estimation," Journal of the American Statistical Association, 106(494):594–607.

Chun H, Zhang X, and Zhao H (2015), "Gene Regulation Network Inference With Joint Sparse Gaussian Graphical Models," Journal of Computational and Graphical Statistics, 24, 954–974. [PubMed: 26858518]

Danaher P, Wang P, and Witten DM (2011), "The joint Graphical Lasso for Inverse Covariance Estimation Across Multiple Classes," arXiv no. 1111.0324.

Finn ES, Shen X, Scheinost D, Rosenberg MD, Huang J, Chun MM, Papademetris X, and Constable RT (2015), "Functional Connectome Fingerprinting: Identifying Individuals Using Patterns of Brain Connectivity," Nature Neuroscience, 18, 1664–71. [PubMed: 26457551]

Fornito A, Zalesky A, and Breakspear M (2013), "Graph Analysis of the Human Connectome: Promise, Progress, and Pitfalls," NeuroImage, 80, 426–444. [PubMed: 23643999]

Fox MD, and Greicius M (2010), "Clinical Applications of Resting State Functional Connectivity," Frontiers in Systems Neuroscience, 4.

Friedman JH, Hastie TJ, and Tibshirani RJ (2008), "Sparse Inverse Covariance Estimation With the Graphical Lasso," Biostatistics, 9, 432–441. [PubMed: 18079126]

Fukumizu K, Bach FR, and Jordan MI (2009), "Kernel Dimension Reduction in Regression," The Annals of Statistics, 37, 1871–1905.

Greene AS, Gao S, Scheinost D, and Constable RT (2018), "Task-induced Brain State Manipulation Improves Prediction of Individual Traits," A Nature Communications, 9, 2807.

Hastie TJ and Tibshirani RJ (1990), Generalized Additive Models, New York: Chapman & Hall.

Hearne LJ, Mattingley JB, and Cocchi L (2016), "Functional Brain Networks Related to Individual Differences in Human Intelligence at Rest," Scientific Reports, 6, 32328. [PubMed: 27561736]

Kolar M, Song L, Ahmed A, and Xing EP (2010), "Estimating Time-Varying Networks," The Annals of Applied Statistics, 4, 94–123.

Lee K-Y, Li B, and Zhao H (2016a), "On an Additive Partial Correlation Operator and Nonparametric Estimation of Graphical Models," Biometrika, 103, 513–530. [PubMed: 29422689]

_____ (2016b), "Variable Selection Via Additive Conditional Independence," Journal of the Royal Statistical Society, Series B, 78, 1037–1055.

Lee W, and Liu Y (2015), "Joint Estimation of Multiple Precision Matrices With Common Structures," The Journal of Machine Learning Research, 16, 1035–1062. [PubMed: 26568704]

Li B, Chun H, and Zhao H (2012), "Sparse Estimation of Conditional Graphical Models With Application to Gene Networks," Journal of the American Statistical Association, 107, 152–167. [PubMed: 24574574]

Li B, and Solea E (2018), "A Nonparametric Graphical Model for Functional Data With Application to Brain Networks Based on fMRI," Journal of the American Statistical Association, 113, 1637–1655.

Liu H, Han F, Yuan M, Lafferty J, and Wasserman L (2012), " High-Dimensional Semiparametric Gaussian Copula Graphical Models," The Annals of Statistics, 40, 2293–2326.

Peng J, Wang P, Zhou N, and Zhu J (2009), "Partial Correlation Estimation by Joint Sparse Regression Models," Journal of the American Statistical Association, 104, 735–746. [PubMed: 19881892]

Qiao X, Guo S, and James GM (2019), "Functional Graphical Models," Journal of the American Statistical Association, 114, 211–222.

Qiu H, Han F, Liu H, and Caffo B (2016), "Joint Estimation of Multiple Graphical Models from High Dimensional Time Series," Journal of the Royal Statistical Society, Series B, 78, 487–504.

Rothman AJ, Bickel PJ, Levina E, and Zhu J (2008), "Sparse Permutation Invariant Covariance Estimation," Electronic Journal of Statistics, 2, 494–515.

Smith R, Lane RD, Alkozei A, Bao J, Smith C, Sanova A, Nettles M, and Killgore WDS (2018), "The Role of Medial Prefrontal Cortex in the Working Memory Maintenance of One's Own Emotional Responses," Science Reports, 8, 3460.

Song M, Zhou Y, Li J, Liu Y, Tian L, Yu C, and Jiang T (2008), "Brain Spontaneous Functional Connectivity and Intelligence," Neuroimage, 41, 1168–1176. [PubMed: 18434203]

Tsay RS, and Pourahmadi M (2017), "Modelling Structured Correlation Matrices," Biometrika, 104, 237–242.

Wang J-L, Chiou J-M, and Muller H-G (2016a), "Functional Data Analysis," Annual Review of Statistics and Its Application, 3, 257–295.

Wang Y, Kang J, Kemmer PB, and Guo Y (2016b), "An Efficient and Reliable Statistical Method for Estimating Functional Connectivity in Large Scale Brain Networks Using Partial Correlation," Frontiers in Neuroscience, 10, 1–17. [PubMed: 26858586]

Wang Y, Yin Y, Sun YW, Zhou Y, Chen X, Ding WN, Wang W, Li W, Xu JR, and Du YS (2015), "Decreased Prefrontal Lobe Interhemispheric Functional Connectivity in Adolescents With Internet Gaming Disorder: A Primary Study Using Resting-State fMRI," PLoS ONE, 10.

Wei Z and Li H (2008), "A Hidden Spatial-Temporal Markov Random Field Model for Network-Based Analysis of Time Course Gene Expression Data," Annals of Applied Statistics, 2, 408–429.

Xu KS, and Hero AO (2014), "Dynamic Stochastic Blockmodels for Time-Evolving Social Networks," IEEE Journal of Selected Topics in Signal Processing, 8, 552–562.

Yuan M, and Lin Y (2007), "Model Selection and Estimation in the Gaussian Graphical Model," Biometrika, 94, 19–35.

Zanto TP, and Gazzaley A (2013), "Fronto-Parietal Network: Flexible Hub of Cognitive Control," Trends in Cognitive Sciences, 17, 602–623. [PubMed: 24129332]

Zhang J, and Cao J (2017), "Finding Common Modules in a Time-Varying Network With Application to the Drosophila Melanogaster Gene Regulation Network," Journal of the American Statistical Association, 112, 994–1008.

Zhang T, Wu J, Li F, Caffo B, and Boatman-Reich D (2015), "A Dynamic Directional Model for Effective Brain Connectivity Using Electrocorticographic (ECoG) Time Series," Journal of the American Statistical Association, 110, 93–106. [PubMed: 25983358]

Zhu Y, and Li L (2018), "Multiple Matrix Gaussian Graphs Estimation," Journal of the Royal Statistical Society, Series B, 80, 927–950.
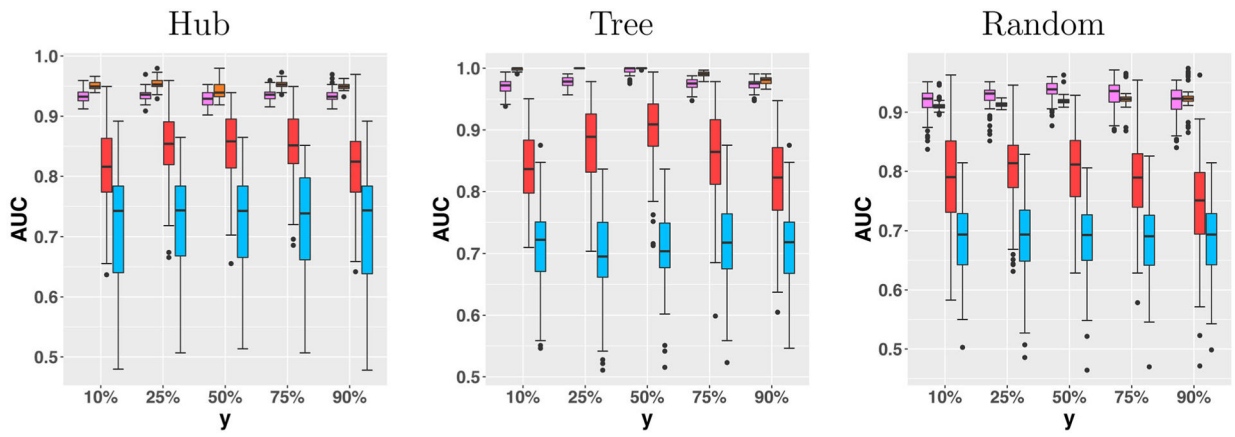
Figure 1.

Area under the ROC curve for the estimated graph with respect to the external variable Y, under three graph structures, and by the four methods. From left to right: CPO, Average, Unconditional, Majority. The graph size $p = 10$.
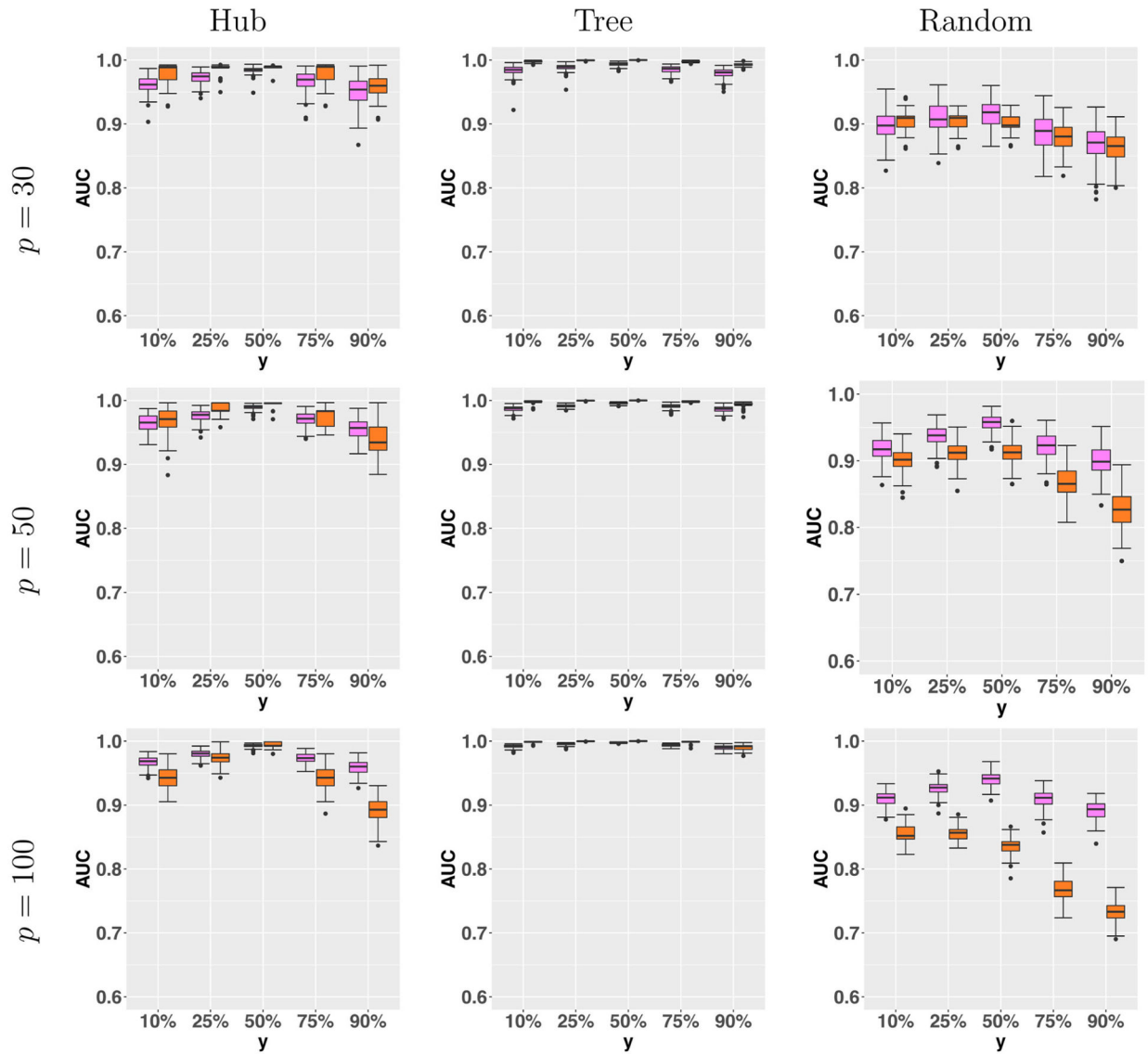
Figure 2.
Area under the ROC curve for the estimated graph with respect to the external variable *Y*, under three graph structures, and by the two methods. From left to right: CPO, Average. The graph size *p* = {30, 50, 100}.
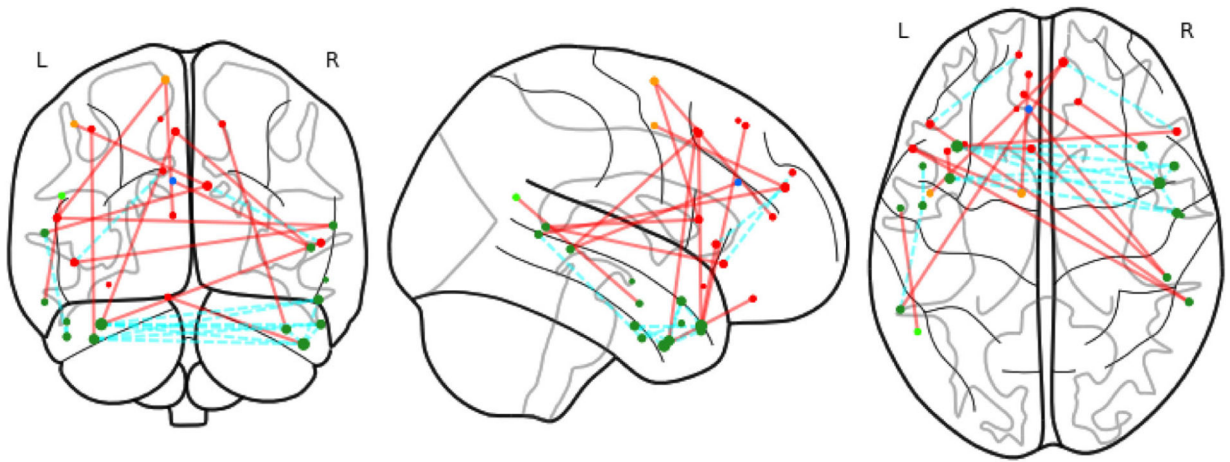
Figure 3.
Medial frontal network, with significant edges. Different colors of nodes indicate the ROIs in different lobes: prefrontal, motor, parietal, temporal, limbic and cerebellum. Red lines indicate inter-lobe edges and cyan ones for inner-lobe edges.
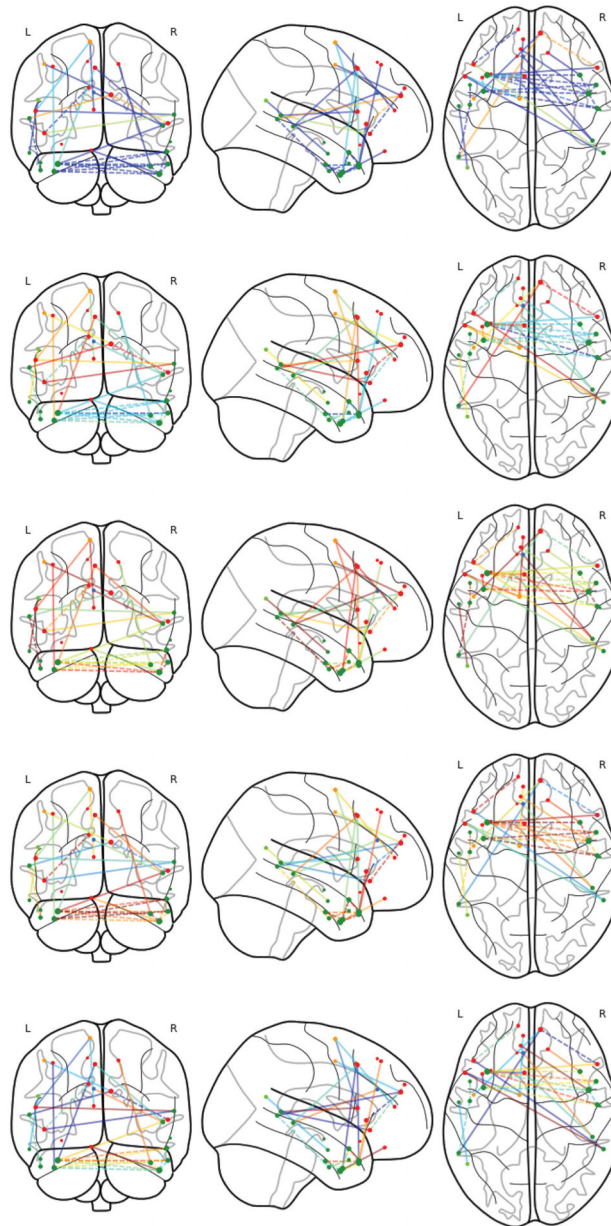
Figure 4.
Medial frontal network changes, with respect to the intelligence score at 7, 11, 15, 19, 23.
Blue color represents the small H-S norm value of CPO, green the medium norm value, and
red the high H-S norm value.

Table 1.

Simulation setup: the edges under three graph structures.